UNIVERSIDAD AUTÓNOMA DE MADRID
FACULTAD DE CIENCIAS
PROGRAMA DE DOCTORADO EN BIOFÍSICA

Underlying Complexity in the Mechanical Unfolding of Fast Folding Proteins as Explored by Atomic Force Microscopy

Memoria para optar el grado de Doctor en Biofísica

Jörg Schönfelder

Thesis Director: Prof. Víctor Muñoz van den Eynde
Co-Director: Dr. Raúl Pérez-Jiménez

MADRID, CURSO ACADÉMICO 20014/2015

La realización de esta tesis doctoral ha sido posible gracias a la concesión de una beca predoctoral del Consejo Superior de Investigaciones Científicas (CSIC) (JAE-Pre011) del Ministerio de Educación y Ciencia de España y de un contrato becario del programa Marie-Curie Excellente Grant.

La tesis ha sido elaborada en el Centro de Investigaciones Biológicas (CIB), en el Centro Nacional de Biotecnología (CNB) y en el Insituto Madrileño de Estudios Avanzados (IMDEA) de Nanociencia en Madrid y en el Centro de Investigación Cooperativa en Nanociencias (CIC nanoGUNE) en San Sebastián.

"..every atom in our bodies was once part of a star"

Gattaca

Acknowledgement

I want to express my special thanks to my thesis supervisor Victor Muñoz, who gave me the opportunity to work in the exciting field of protein folding and for his support and for fruitful discussions during all this years and especially during the last period of my thesis. I'm very grateful for the possibility of traveling to international conferences and workshops, where I had the chance to meet other scientists, present and discuss my research. Furthermore for giving me the opportunity of having a short two month long stay as a guest researcher in the laboratory of Julio Fernández at the Columbia University in New York, USA and for further four stays during the last two years of my thesis in the laboratory of Raúl Pérez-Jiménez at the CIC nanoGUNE institute in San Sebastián, Spain. Without this support the performed work would have not been possible.

Special thanks to Raúl Pérez-Jiménez, who was teaching me how to use the force-clamp AFM during my stay at the Columbia University and who later became my co-supervisor of my thesis as I did the main part of my research in his laboratory in San Sebastián. Thank you a lot Raúl for your patience, for your shared research experience and for giving me this opportunity to use your AFM instrument as much as possible during my stays at your laboratory.

I also want to thank Mariano Carrión-Vázquez for providing the titin I27 domain, the ubiquitin pFS and the expression vector pRSETa vector, which I needed to start the measurements.

Furthermore this work would not have been possible without all the help and support from my colleagues over the years from Victor´s and Eva de Alba's lab who I want to mention here. Special thanks to you Javier. Thank you for your willingness, encouragement and patience to teach me how to interpret the AFM data and especially how to build, express and purify the highly sophisticated polyprotein samples at its best. Unquestionable the big majority of the results of my research would not have been discovered without your help. Thank you Luis for your guidance, shared experience and fruitful discussions especially at the very beginning of my thesis, for me as a physicist your help in the molecular biology

part of my research was really needed in order for me to do the work independently afterwards. Thank you Ravi, Abhinav and Rajendra you were always patient and very helpful in solving and discussing computational issues, which resulted also in an important part of my research. Many thanks to you Mila not only for your help during the last years, especially during the part of writing up the thesis, your encouragement in the proof-readings and discussions were a very important factor in order to put this thesis into shape.
Finally I thank all of you Agnieszka, Athi, Clara, Celia, David, Eduardo, Evelyn, Lorenzo, Maggie, Malwina, Marta, Michele, Mourad, Nacho, Nicola, Pacho, Patricia, Pilar, Satish, Susana, Tanay for contributing to create a very stimulating and comfortable environment in the lab, which was very needed to carry out my research in the best possible way.
Furthermore I thank to the people who I met during my stays at the Columbia University to Jorge, Sergi, Rodolfo, Ronen, Pallav, and Ionel and at the nanoGUNE to Aitor, Alvaro, Marie, Nerea and Simon for their support.
I want to thank Anna for all her support and motivation over the last years and especially for her encouragement during the last period of writing the thesis.
Finally I would like to thank my parents, my brother, my sister and Peter for their support and their patience during all these years, as without their encouragement, this work would not have been possible.

Contents

Resumen

Las proteínas son polímeros flexibles y, al mismo tiempo, macromoléculas muy sofisticadas, capaces de llevar a cabo una gran variedad de funciones biológicas. Entre los numerosos ejemplos podemos destacar su capacidad de regular los procesos fisiológicos en células vivas en forma de motores moleculares, realizar funciones de transporte mediante canales intramembrana, regular caminos de señalización durante la transcripción y la transducción de la información genética –las llamadas proteínas de *scaffold*–, o su papel crucial en múltiples formas durante la compleja cascada de eventos que tiene lugar durante la respuesta inmune. El buen funcionamiento de las proteínas subyace principalmente en dos propiedades: primero, en su habilidad para formar estructuras tridimensionales codificadas en su secuencia de aminoácidos; segundo, en la posibilidad de dicha estructura a cambiar debido a estímulos externos tales como variaciones en la temperatura, en la presencia de agentes químicos, o en la aplicación de fuerzas mecánicas. Descifrar las reglas que regulan el correcto plegamiento de las proteínas en su estructura tridimensional funcional es un problema de gran interés en muchos ámbitos de investigación, como ciencia básica, biotecnología, ciencia de materiales, y medicina, ya que haría posible el diseño de nuevas proteínas con funciones específicas deseadas. Aún así, a pesar de la gran dedicación de la comunidad científica y de las extraordinarias mejoras experimentales durante las últimas cinco décadas, ni la pregunta de cómo se pliegan las proteínas ha sido respondida ni es posible su diseño acorde a una función. En este sentido, el estudio experimental del plegamiento de las proteínas está limitado por la dificultad de medir con la necesaria elevada resolución temporal, estructural, y a nivel de molécula individual, para poder observar con detalle el mecanismo que está detrás del comportamiento aparentemente simple que se obtiene con métodos experimentales convencionales.

El reto científico actual para estudiar el plegamiento molecular es el desarrollo de nuevas tecnologías que permitan caracterizar las múltiples configuraciones que una proteína explora antes de llegar a su correcta estructura (lo que se conoce como el "camino de plegamiento"). A pesar de todo, la mayoría de proteínas investigadas hasta el presente

sugieren un comportamiento de dos estados (plegado-desplegado), y la caracterización de la complejidad cinética durante el plegamiento es todavía un desafío. Incluso la recién aplicación de técnicas que permiten seguir el comportamiento de una única molécula no ha conseguido caracterizar en detalle el plegamiento y en muchos casos los resultados experimentales permiten solo observar un proceso de dos estados, lo que hace muy difícil la validación de predicciones mecanísticas obtenidas por la teoría y las simulaciones computacionales. Recientemente se han desarrollado técnicas de espectroscopía de fuerza de molécula individual (EFMI), que permiten aislar una única proteína y aplicarle fuerzas del orden de picoNewtons, favoreciendo su desplegamiento mecánico. Dichas técnicas posibilitan caracterizar los procesos de plegamiento y desplegamiento de una única proteína a nivel individual mediante cambios que tienen lugar en una coordenada de reacción bien definida, como es la extensión molecular de la proteína. Así pues, mediante el uso de la EFMI se puede caracterizar el efecto de una fuerza externa en el paisaje de energía libre de la proteína en función de la extensión molecular.

En esta tesis doctoral se realiza EFMI mediante el uso de la microscopía de fuerza atómica para caracterizar el desplegamiento mecánico de distintas proteínas que exhiben una cinética de plegamiento ultrarápida, y cuyo plegamiento ha sido hasta ahora estudiado en experimentos de desnaturalización química y térmica. El objetivo es determinar si, a pesar de mostrar un comportamiento muy simple durante su plegamiento, estas proteínas contienen la complejidad que la teoría del paisaje de energía libre y que simulaciones computacionales predicen. En particular, en esta tesis se investigan tres proteínas con diferentes mecanismos de plegamiento: la proteína Csp (proteína B cold shock de *Thermotoga Maritima*) se pliega en aproximadamente 1 milisegundo siguiendo un proceso de dos estados, la proteína BBL (de *E. Coli*) en microsegundos mediante un único estado (*downhill*), y la proteína gpW (de *bacteriophage* λ) que sigue un mecanismo de plegamiento de cerca *downhill*. Estas tres proteínas ultra-rápidas se caracterizan experimentalmente usando microscopio de fuerza atómica (MFA) y a través de simulaciones de dinámica molecular dirigida (DMD).

Para realizar EFMI, una parte muy importante del proyecto experimental ha requerido el diseño de poliproteínas mediante técnicas de manipulación de AND, expresión y purificación por HPLC. Dichas poliproteínas están formadas por la proteína de estudio y varias copias otras proteínas, como la titina I27 o la ubicuitina, cuyo patrón de plegamiento y desplegamiento ha sido muy estudiado con MFA. Así, cuando en nuestros experimentos de EFMI reconoces el conocido patrón de desplegamiento/plegamiento de las poliporteínas de

titina I27 o ubicuitina, sabes sin lugar a dudas que también la fuerza ejercida por el MFA también actua sobre la proteína de estudio (Csp, BBL o gpW). En el proyecto, se han probado distintos diseños de poliporteínas con el objetivo de mejorar la señal obtenida por la proteína de estudio. Los experimentos con MFA también han requerido la optimización en sus modos operacionales (fuerza constante y velocidad constante controlada en fuerza) para mejorar la detección del desplegamiento de la proteína de estudio. En paralelo, las simulaciones de DMD se han llevado a cabo con el programa NAMD i un campo de fuerzas CHARMM22 as a para complementar los resultados experimentales. Finalmente, también se han realizado modelos estocásticos de la cinética de las proteínas para dar una interpretación rigurosa a los resultados experimentales. Los resultados obtenidos a lo largo de la tesis para las tres proteínas Csp, BBl y gpW indican claramente la gran complejidad en el mecanismo de desplegamiento molecular. En el caso de la proteína Csp, se ha observado en experimentos de MFA realizados con los modos de fuerza constante o a velocidad constante que el proceso de desplegamiento a fuerzas bajas (entre 20 y 80 pN) á regulado por la presencia de múltiples estados intermedios en diferentes caminos de reacción. Además, ha sido posible modular la variedad de caminos de desplegamiento mediante la acción una la fuerza mecánica en los extremos de la proteína. Este peculiar resultado no se ha observado en experimentos de EFMI realizados anteriormente, y es por lo tanto uno de los resultados más importantes de esta tesis. Las simulaciones de DMD muestran un acuerdo excelente con las medidas experimentales. Los resultados para las proteínas gpW y BBL muestran que la combinación del mecanismo de plegamiento tipo *downhill* junto con una velocidad de plegamiento del orden de microsegundos puede provocar una gran heterogeneidad en los patrones de desplegamiento medidos entre 1 y 20 pN. En particular, para la proteína gpW los experimentos a velocidad constante se realizaron a 1 pN/s (muy cerca del equilibrio) y los experimentos a fuerza constate se realizaron a 5 pN. Hasta ahora, este rango de fuerzas solo se había explorado en experimentos de EFMI realizados con pinzas ópticas o magnéticas, de modo que han requerido de la mejora sustancial de la capacidad de medida del MFA. Para el caso de la proteína gpW, el uso de modelos estocásticos ha resultado imprescindible para explicar la gran variedad de patrones de desplegamiento.

En conclusión, durante esta tesis doctoral se ha mejorado la capacidad del MFA para detectar la gran complejidad de caminos de desplegamiento en proteínas ultra-rápidas en experimentos de EFMI. Así, el uso del MFA para medir el desplegamiento mecánico de proteínas con diferentes mecanismos de plegamiento resulta un enfoque nuevo, muy capaz de descubrir nueva fenomenología en el estudio del plegamiento molecular.

Abstract

Proteins are both flexible polymers and sophisticated biological macromolecules capable of performing an enormous variety of biological functions enabling them to manage physiological tasks of molecular motors in the living cell, carriers, scaffolds, transport channels or e.g tasks during the highly complex cascade reaction of the immune defense. All these functions rely on two remarkable capacities: first the ability of proteins to self-assemble (fold) into 3D structures encoded into their specific amino-acid sequence and second its change of shape in response to an external stimuli like temperature, chemicals or mechanical force.

Decoding the rules and mechanisms by which proteins fold would have a very deep impact into many fields of research like basic science, biotechnology, biomaterial science and medicine as it would allow designing proteins with new desired functionalities. However, regardless of tremendous scientific efforts as well as experimentally improvements during the last five decades, neither the question of how proteins fold is solved nor the design of new proteins is being possible. One of the most restricting limitations in this sense has been the difficulty of studying the protein folding process with the necessary high temporal, structural, and single-molecule resolution in order to really probe the folding mechanisms underlying the surprisingly simple experimental observations of conventionally performed protein folding experiments.

Therefore the major challenge which motivates research in protein folding has been the development of technologies for resolving the multiple unfolding pathways and other detailed features of folding mechanisms that lie under the simple characterized two-state folding behavior. However detecting such kinetic complexity has still remained desirable and the majority of proteins investigated so far are suggested to un(fold) via a simple two-state mechanism. Such apparent simplicity restricts probing the folding mechanism of the protein in detail, and makes it difficult to cross validate mechanistic predictions from theory and computer simulations. Even the recent application of single-molecule experimental techniques to investigate two-state folding has not directly resolved further details.

The recently developed single-molecule force spectroscopy (SMFS) technique, which applies a mechanical force in the pN range to grab and pull from single protein molecules in order to provoke their mechanical unfolding enables the characterization of individual unfolding behaviors. This permits the investigation of a well defined single reaction coordinate (the end to end distance of the protein) and their effect on the free energy landscape of the protein.

This research project applies SMFS using the atomic force microscope (AFM) to characterize the mechanical unfolding of several proteins that exhibit fast to ultrafast folding kinetics, which are determined from isotropic temperature and chemically-induced denaturation experiments. The goal is to determine whether the relatively simple folding behavior displayed by these proteins is in fact hiding an underlying complexity that energy landscape theory and computer simulations predict for these proteins.

In particular, within this project three fast folding proteins with different folding mechanisms are investigated that fold at rates that go from the moderately fast (1ms) two-state folder Csp (Cold shock protein B from *Thermotoga Maritima*) to the microsecond rates of the one-state folder BBL (from *E.coli*) and the nearly downhill folder gpW (from bacteriophage λ). These three small fast folding proteins are characterized both experimentally using the AFM and computationally using Steered Molecular Dynamics (SMD) simulation.

In order to perform the SMFS experiments a huge part of the project involved the design by DNA manipulating techniques, expression and production by HPLC purification of so-called polyprotein samples that are needed for the mechanical manipulation by the AFM. This implies introducing the case study protein in a polyprotein composed of several copies of the well studied mechanical force bearing titin I27 domain or ubiquitin protein, both being proteins that are used as a molecular fingerprint in the AFM studies. The basic idea of the molecular fingerprint is that, when the characteristic mechanical unfolding pattern of the titin I27 domain or ubiquitin protein is observed during the measurements it can be undoubtedly guaranteed that a mechanical force was applied on the polyprotein sample and therefore also on the inserted protein of choice. During this project several attempts of producing a polyprotein construct were tried in order to improve the unfolding signal of the single copy of the protein of study.

The AFM experiments involved a huge effort of optimization and improvement in both constant velocity and force clamp mode in order to detect the unfolding pattern of the case study protein. SMD simulations were performed in parallel with the help of an in-house HPC cluster using the NAMD software and a CHARMM22 force field as a further com-

plement to the experimental observed behavior. Part of the experimental data analysis required the elaboration of a stochastic kinetic model that enabled the simulation of the experimental observations, which therefore was very helpful in interpreting the experimental data.

Results for all three proteins Csp, BBL and gpW undoubtedly reveal an underlying complexity of in the mechanical unfolding mechanism.

Csp was found to unfold through multiple intermediate states and via many different pathways during force ramp and constant force AFM experiments in the low force regime between 20 and 80pN. Furthermore mechanical force was found to be able to modulate the variety of existing unfolding pathways of Csp. This very particular behavior of a protein under force was not reported before by any known SMFS study and is therefore one of the main discoveries of this thesis. Furthermore in the case of Csp, SMD simulations show an excellent agreement to the experimental observed behavior.

The result of gpW and BBL revealed that the property of the fast folding rate in microseconds in combination with the downhill folding mechanism can provoke the discovered heterogeneity in unfolding patterns detected in a range of 1-20pN. With the gpW protein, force ramp measurements were able to be performed at a rate of 1pN/s (near equilibrium) and at a constant force of 5pN. This performance was therefore pushing the measuring capacity of the force-clamp AFM to a before unknown low force regime, which was thought to be only accessible for SMFS experiments using optical or magnetic tweezers. Additionally the employed stochastic kinetic model was able to explain high variety of observed unfolding patterns of gpW.

In sum, the significance of this thesis is that it unveils the capacity of SMFS by using the AFM for detecting a high underlying complexity in unfolding patterns of fast folding proteins. Thus applying the AFM technique to fast folding proteins following different types of folding mechanisms indeed is a further important approach to give insights into the field of protein folding.

List of Abbrevations

aa:	amino acid
AFM:	Atomic Force Microscope
Csp:	Cold Shock Protein
DNA:	Deoxyribonucleic acid
HPLC:	High performance liquid chromatography
FEL:	Free energy landscape
MD:	Molecular Dynamics
MFPT:	Mean First Passage Time
NAMD:	NAnocsale Molecular Dynamics
RNA:	Ribonucleic Acid
PCR:	Polymerase Chain Reaction
SMD:	Steered Molecular Dynamics
SMFS:	Single Molecule Force Spectroscopy
SSA:	Stochastic Simulation Algorithm
TS:	Transition State
VMD:	Visual Molecular Dynamics
WLC:	Worm-Like-Chain

Part I

Introduction

Chapter 1

Protein folding

In the following chapter terms and definitions to describe the protein structure. Then the wide field of protein folding will be introduced until a minimal set of important concepts, which are needed to put the whole thesis into context.

1.1 Introduction to proteins

Proteins are by far the most structurally complex and sophisticated biological macromolecules known in the living cell, which constitute most of the cell's dry mass. They are essential parts of the cell, necessary to form the cell structure and to execute all the cell's functions. Therefore an enormous variety of protein functions are found. For example proteins are found in the cell membrane, which act as channels or pumps that control the transfer or exchange of ions or small molecules from the outside of the cell to the inner cytoplasm. Others can act as carriers to communicate between cells. Yet others exist as a type of molecular motors which can change its conformation in order to fulfill its task. The spectrum of specialized protein ranges from acting as antibodies, toxins, hormones, elastic fibers to the source of luminescence and many other not listed functions are executed.
This remarkable versatility has been developed and fine-tuned over billions of years of evolutionary history of protein structure and chemistry. The understanding of protein structure and self-assembly (folding) has a deep impact in into many fields of research as it would allow to design proteins with a desired and therefore specialized function (nanomachines) and will be explained in the two following sections.

1.2 Protein structure

Proteins are composed of discrete subunits known as the amino acids (aa). There are 20 different amino acids existing in natural proteins, each with different chemical properties (see figure 1.1).

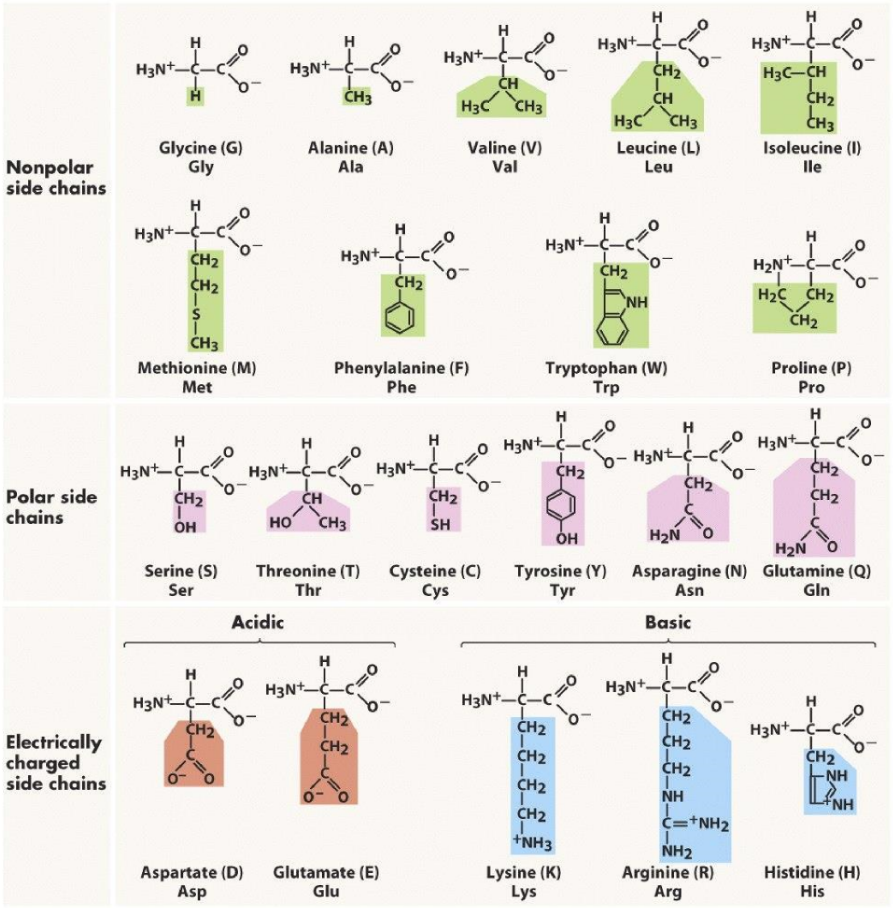

Figure 1.1: The protogenic amino acids

A protein molecule is formed by a long chain of these amino acids, each linked to the next one by a covalent so-called peptide bond. This bond is between the carboxyl group of the first aa and the amino group of following aa. Within the polypeptide chain the amino acids are also called residues. The backbone or main chain of a protein always begins with an amino (N) terminus and ends with a carboxyl (C) terminus.

Each aa consist of a central carbon atom (called C_α) and a hydrogen atom attached to it. Furthermore a carboxyl group (COOH) and amino group (NH_2) is connected to the C_α atom. Therefore amino acids are identical building blocks just different by the side chain group which occupies the forth binding possibility of the C_α atom. Hence only the side chain provides the specific property of the aa, which can be nonpolar, polar or negatively/positively charged (see again figure 1.1)).

The final structure and properties of any protein is given by the the exact location of each amino acid in the long polypeptide chain sequence (primary structure) and depending on the corresponding involved side chains, the protein results more flexible or more rigid (Figure 1.2). There are two dihedral angles, named psi (ψ) around the amino group and phi (ϕ) around the carboxyl group respectively bond with the C_α atom, which offer a certain degree of freedom and therefore rotational mobility.

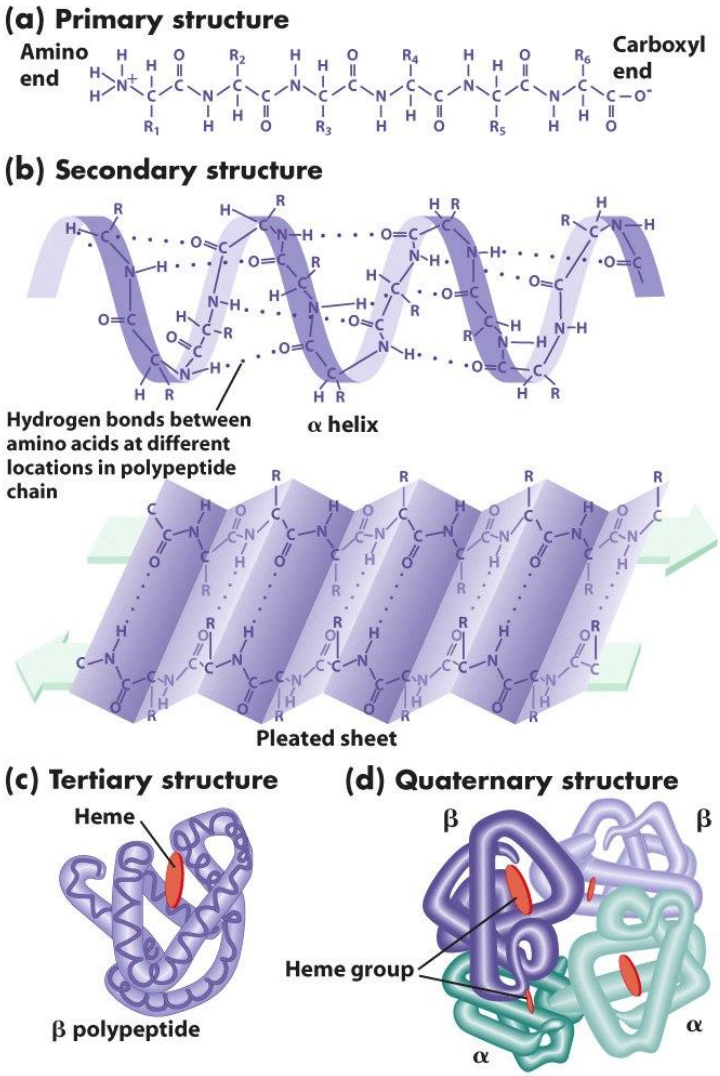

Figure 1.2: The protein structure

Inter- and intra-peptide forces are now causing the protein to find and to maintain its 3D structure. These forces mainly involve weak interactions like hydrogen bonding, electrostatic, van der Waals and hydrophobic interactions, which gives every protein a certain mobility additional to the amino acid composition.

Beside the resulting enormous variety of protein structures which can evolve from the primary structure of the protein, there are two common structural motifs namely the α-

helix and the β-sheet made by two β-strands. They form the so-called secondary structure of a protein, which neutralizes the polar groups of the backbone by forming hydrogen bonds. Whereas the α-helix is more flexible the β-sheet is more rigid. The third common structural motif is the so-called loop region, which connects for example several α-helices or two β-strands to form a β-sheet within the protein. Loop regions are commonly less structured and therefore highly flexible.

Finally, the spatial interaction between the secondary structure motifs forms the so-called tertiary structure of the protein. This spatial interaction is mainly caused by hydrophobic interaction, disulphide bonds and salt bridges. Most proteins need to adopt the tertiary structure to be biologically active.

Additionally, when tertiary structure of several protein interact with each other a so-called quaternary-structure is formed.

1.3 Historical introduction to protein folding

After discovering how proteins are produced within the cell, the question of how proteins fold from the simple amino acid primary structure into the final tertiary structure became more and more important. One of the first milestones in protein science was the finding of reversibility of totally denatured (unfolded) proteins and the postulation of the thermodynamic hypothesis by Christan Anfinsen and colleagues [1, 2] in the 1960's. From there on it was clear that the native protein structure is the thermodynamic stable structure, which depends just on the protein amino acid sequence and on the conditions of the surrounding solution. This finding enabled to investigate protein folding in the test tube. It also states that although evolution can change the amino acid sequence, the principles of folding itself are determined by principles of physical chemistry. Since then enormous experimental, computational and theoretical progress has been made in this field [3, 4]. However many important open questions still remain today.

Among others there are three important unsolved issues connected to protein folding [5]:

1. What are the thermodynamics behind the tertiary protein structure? I.e. what determines the 3D protein structure for a given amino acid sequence and protein environment?

2. How does the protein find its folding pathway so fast? Are there existing intermediate states along this folding pathway?

3. How can the tertiary protein structure be predicted from the primary structure?

1.4 Theoretical and computational models of protein folding

1.4.1 The energy landscape

Theoretical models to answer the protein folding problem were driven by the Cyrus Levinthal´s argument made in the late 1960´s, which became later the Levinthal´s paradox [6]. Levinthal argued that when each bond connecting amino acids has three possible states, the overall folding process for a 100aa containing protein would take around 10^{27}years, when the protein searches randomly for its tertiary structure. As proteins are known to fold in a time scale from seconds down to microseconds, specific folding pathways may allow the protein to find its native state in the proper amount of time.
With the theoretical description of a possible folding behavior through a discrete pathway, experimental groups like Tsong et al [7] and Ikai and Tanford [8] were convinced that a sequential folding pathway could be detected experimentally. After finding intermediate states, several pathway models were used to describe the experimental results. E.g models like the so-called on-pathway [7], in which the protein find its native state with or without existing intermediate states and the off-pathway [8], in which the protein can misfold or be kinetically trapped in a state different to the native state. As the experiments at this time just probed the average behavior of the protein coming from a single/multi-exponential decay of optical properties that monitor changes in the protein structure, these models were sufficiently well describing the experimental results.
However two developments changed this view of modeling protein folding with a discrete pathway model. First due to enormous experimental advances to study protein kinetics down to the atomic level especially with NMR spectroscopy [9] and the new experimental possibilities to characterize the kinetics of fast two-state folding proteins [10]. And secondly due to the theoretical advances in using statistical mechanics models with highly simplified representations of protein chain geometries and interactions. With this so-called lattice-models the free energy surface on which the protein folds has been described with a minimum of parameters and much could be explained like the relative balance between conformational entropy and stabilization energy or the role of cooperativity in the folding

process [11, 12].

Hence the pathway-model, in which protein folding consists of a clear sequential connection between macroscopic states (sequentially folding) was soon replaced by the energy landscape theory [13, 14, 15], which describes protein folding as a diffusion-like process on the its energy landscape driven by Brownian motion (parallel folding). I.e proteins are described to fold from a conformational ensemble with high conformational entropy downwards to a fewer populated ensemble of conformations of minimum stabilization energy (the native state) through a stabilization energy gradient. This theory is based on statistical mechanics models of glass and phase transitions [16]. As the folding of a random chosen heteropolymer shows highly complex multi-exponential kinetics more associated with the kinetics of glasses, native proteins must have evolved to avoid this kinetic complexity. Therefore while the energy landscape of a heteropolymer with a randomly chosen sequence of amino acids is rugged, the energy landscape of a native protein has to be smoothed to form a funnel.

Often these funnels are sketched in three or two dimensions to visualize the basic concept of their idea (see figure 1.3,1.4).

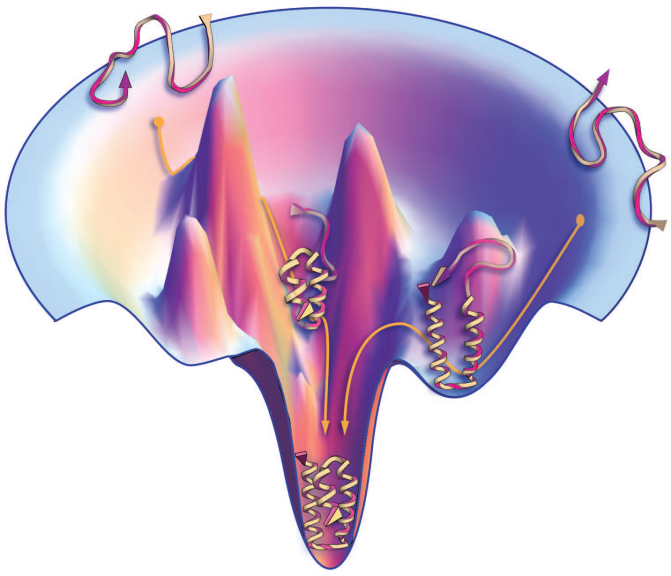

Figure 1.3: Funnel shaped energy landscape [4]

The vertical axes represents the internal free energy of a given chain configuration and the width of the funnel represents the conformational entropy. Barriers to folding (around k_BT) are relatively small compared to the total binding energy (around 100 k_BT) of the native folded structure, simple because during the protein folding process no covalent bonds need to be broken (see Figure 1.4) [17].

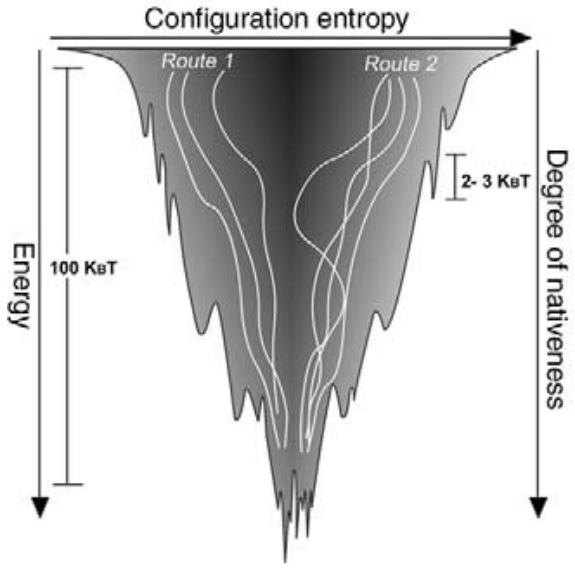

Figure 1.4: 2D view of the funnel shaped energy landscape [17]

The overall process of protein folding starts from the denatured state, in which the protein populates a large ensemble of structures. The polipeptide chain now can fold by numerous parallel pathways, potentially adopting multiple partially folded ensembles until the native state ensemble of minimum stabilization energy. As it has been mentioned before this general organization is not typical for all polymers containing amino-acids. It is more a result of evolution. Evolution achieved to select amino-acid sequences in which the interactions within the functional 3D protein structure are not in conflict instead they are mutually supportive (i.e. a single mutation cannot destroy the overall functional protein structure) and cooperatively leading to a low-energy structure. The interactions, which favor the protein folding are called 'minimally frustrated'. For comparison the folding of a randomly chosen heteropolymer will be 'highly frustrated', meaning that the optimal lowest free-energy cannot be found due to conformational restrictions.

1.4.2 Protein Folding as a projection of the energy landscape

Studying protein folding experimentally in vitro consists traditionally in changing the solution conditions from ones that stabilize the unfolded state to the ones that stabilize the folded state and vice versa. E.g. with external perturbation agents like chemical denaturant (Urea or GdmCl) or temperature the protein stability can be tilted out of equilibrium.

To have an idea of how thermodynamic and kinetic parameters can be derived extracted from the protein folding mechanism it is helpful to review a simple chemical reaction. Here the Arrhenius formula can be used, which describes the dependence of the reaction rate k

from the temperature T [18].

$$k(T) = Ae^{-\frac{E_A}{R \cdot T}} \tag{1.1}$$

with R being the gas constant and A, the pre-exponential factor (also called Arrhenius constant), which normally is also temperature dependent, but weakly compared to the reaction rate (negligible). The activation energy E_A is the temperature independent change in energy converting a chemical reaction from reactant state A to the product state B. Here the Arrhenius constant A and the activation energy E_A can be estimated straight forward from a logarithmic plot k(T) vs 1/T. Hence it becomes possible to estimate the activation barrier and/or to observe transition states.

The Arrhenius description was then further developed by Eyring [19], who used the concept of the potential of mean force. Later Kramers [20, 21] added the fact that the motion of chemical reactants obey Brownian motion. Kramers theory assumes that the dynamics of a chemical reaction can be described by one-dimensional diffusion along the reaction coordinate in which both the reactant/product well (A,B) and barrier top need to have parabolic shape. The theory of Kramers gives the rates of transitions between the reactant state A and and product state B as follows:

$$k_{A\to B} = \frac{\omega_A \omega_{TS}}{2\pi\gamma/m} e^{-\frac{\Delta G^0_{TS}}{k_B T}}, k_{B\to A} = \frac{\omega_B \omega_{TS}}{2\pi\gamma/m} e^{-\frac{(\Delta G^0_{TS} - \Delta G^0_B)}{k_B T}} \tag{1.2}$$

The rates depend exponentially on the free activation energy of the states A and B, which is here the difference between the free energy of the transition state TS and the free energy of the states A and B. The pre-exponential factors in equation 1.2 depend on the diffusion of the molecule from states A and B to the transition state TS respectively. Here the characteristic frequencies $\omega_{A/B}$ of the harmonic well A/B influences both rates $k_{A\to B}$ and $k_{B\to A}$. In a similar meaning, ω_{TS} influences the rates of passing over the transition state once it has been reached either forward from state A or backwards from state B. Furthermore the rates will be damped, described by the ratio of the friction coefficient γ experienced by the molecule to its mass m (γ/m). The characteristic frequencies depend furthermore on the shape or curvature of the harmonic well at each state (A,B and TS).

The process of protein folding can be seen of something in between a single chemical reaction and a phase transition of a multi-particle system [16]. As the protein folding can involve several thousands coordinates, it is clearly becoming much more complex than the described single chemical reaction. For example as a consequence of the high conformational freedom of a polipeptide chain, entropy plays a much more important role

in protein folding, than in most small molecules reactions. Hence within the free energy of protein folding has to be considered as an interplay between entropic and energetic terms. Because it is far to complex to consider all possible coordinates explicitly, it is necessary to introduce a reduced set of reaction coordinates. These coordinates characterize the folding reaction and then the full configuration space is described in terms of the chosen coordinates. For example in the case of a simpler chemical reaction, the temperature can be used as a reaction coordinate (see Arrhenius formula 1.1). Or for example if the chemical reaction involves a formation of a bond between two atoms, the distance of the two atoms would be an appropriate reaction coordinate.

However, as it has been mentioned before, finding an appropriate single reaction coordinate in the case of protein folding is much more difficult. A useful approach is to employ one or more progress variables, which then describe the similarity of each conformation to the native state. In this way the energy landscape of a protein will be projected down to a low dimensional system, which includes a small number of order parameters describing the thermodynamic and kinetic properties of the protein folding process. Useful reaction coordinates describing the protein folding process are the radius of gyration R_g, the number of native contacts (Q), the total number of contacts (C) (native and non-native) or the number of incorrect oriented residues (S) [22, 23]. As it will be shown in the following chapter the end-to-end distance of a protein when exposed to a mechanical force is commonly used as a reaction coordinate in force spectroscopy experiments. The protein dynamics are then described by one-dimensional diffusion along the reaction coordinate within an equation similar to Kramers equation 1.2.

1.4.3 Simulating protein folding energy landscapes

As it has been mentioned before lattice models were used to describe the protein folding process on a low dimensional energy surface [11, 12]. These simplified models are then able to perform a large number of simulations required for a meaningful description of the thermodynamics and kinetics of the protein folding reaction.

Within this class of models, the protein is presented as a string (necklace) of beads positioned at sites on a lattice. The beads are then interacting to each other through pairwise contact potentials in that they represent residue-residue interaction in the presence of the equilibrated solvent. Potentials are normally chosen to be attractive in order to approximate the highly compact organized structure of a native protein.

The motions of the polypeptide chain within this model are then simulated with dynamic

Monte Carlo simulations. Therein, small changes in conformation are made repeatably and accepted or rejected regarding to the change in energy. This will be done through an algorithm, in which the probability is higher for a conformation with lower energy than with higher in order to mimic the real protein folding situation. Therefore the folding reaction is directed by the potential energy function.

For a better explanation, in the following example a lattice folding simulation of a 27-mer using two employed progress variables (Q and C) will be given (see figure 1.5).

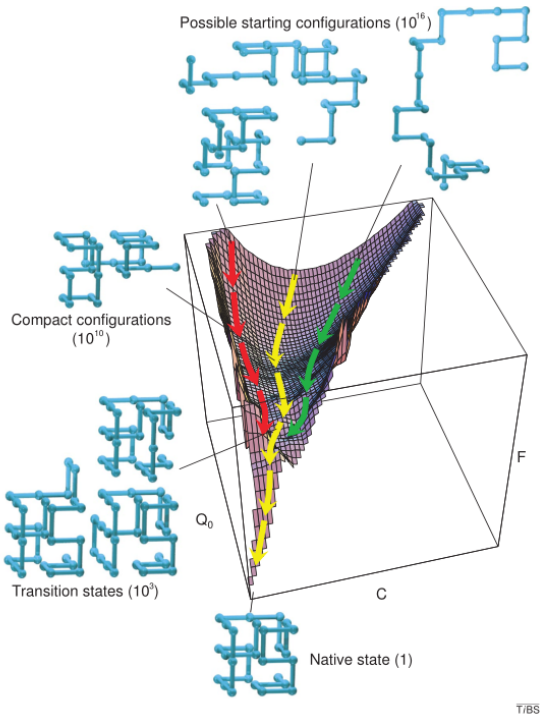

Figure 1.5: Free Energy surface of a 27-mer as a function of native contacts (Q) and the total number of contacts (C) obtained by sampling the accessible configuration space with Monte Carlo Simulation. Taken from [24]

A typical simulation starts within a random configuration, which is similar to the one experimentally obtained after diluting the chemical denaturant to provoke protein folding. On average, the 27-mer chain collapses rapidly into an ensemble with around 60% of all the number of contacts C, of which only 25% are native like (Q). This stage is usually called the collapsed state or molten globule with 10^{10} possible configurations, as it forms a broad broad minimum on the free energy surface. After the collapse, the 27-mer chain encounters the rate limited stage of the reaction, in which the semi-compact conformations are searching for a transition state that leads to the native state.

For the 27-mer chain the transition state ensemble is broad containing 10^3 possible conformations, which incorporates 80-90% native like contacts (Q). Here this simple lattice

model shows that starting from the collapsed state, the 27-mer chain just need to find 1 out of 10^7 possible conformations than 1 out of 10^{16}, which are the possible starting configurations. Derived from this simulation the protein just need to find a small fraction of all of its possible conformations. Hence the Levinthal paradox has been conceptually resolved.

As it can be seen in figure 1.5, the surface of the free energy surface is consists of barriers and local minima. These barriers and local minima are provoked through the interplay of conformational entropy cost and the energy. As there is the direction from high to low energy, which favours a compact protein conformation there exist the contrary direction driven by conformational entropy cost which favors the open protein conformation. Therefore at each conformational state on the free energy surface the balance between entropy and energy drives the state of the protein conformation.

For example, if the contribution to the free energy of the configurational entropy decreases faster than the average energy, a 'bottleneck' results, provoking a barrier on the free-energy surface, which corresponds to the transition state in figure 1.5.

In sum, lattice models have shown that a protein does no need to find all of its possible conformations to fold, because its energy surface restricts the folding search and that there are many transition states existing.

1.4.4 Protein folding predictions from the energy landscape

When the free energy projection produces a good reaction coordinate then the folding kinetics are described as diffusion on that free energy surface. Figure 1.6 shows the projection of the two opposing forces in folding: conformational entropy (in energy units) and stabilization free energy. A projection of the free energy surface onto a single order parameter then represents the degree of native structure in terms of local conformation.

As discussed in the sections before, projecting a high dimensional free energy surface to some or a single reaction coordinate enables to predict the height of the folding/unfolding barrier. A folding barrier arises because of non-perfect compensation of the decrease in stabilization energy and parallel increase of conformational entropy. The exact compensation defines the barrier height and the relative folding rate (one dimensional diffusion along the reaction coordinate). This also states that the free energy surface (the protein folding funnel) implies a general correlation between folding stability and folding rates. Under native conditions the compensation between these two factors typically results in a small free energy barrier that for most small proteins is below 16RT(40 kJ mol^{-1})

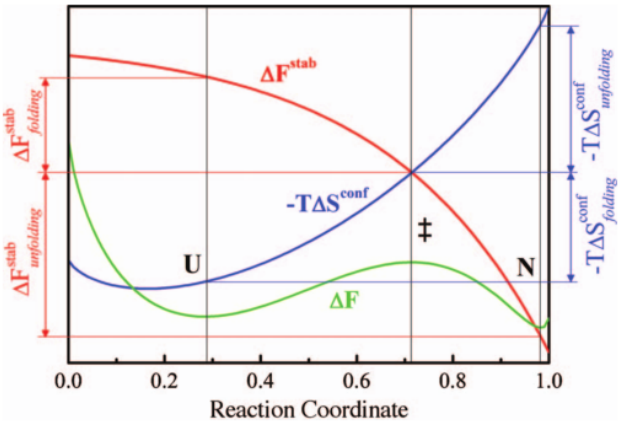

Figure 1.6: Representation of the balance between the two opposing forces in protein folding in a one-dimensional free energy surface. The stabilization free energy is shown in red, the cost in conformational entropy in blue, and the resulting free energy surface in green. The partitioning of stabilization free energy ΔF^{stab} and conformational entropy ΔS^{conf} on both sides of the barrier top is shown on the sides of the figure (see [25]).

Bryngelsson et al [13] defined this case of protein folding as Type I or two-state scenario, in which a barrier separates the denatured and native state. It proposes a two-state transition as an equilibrium between a single folded conformation and an unfolded state with no accumulation of stable intermediates. Herein the population is always separated between folded and unfolded state as the population of the conformational ensemble on the top of the barrier height (transition state) is negligible. Such a 'bottleneck' blocks a high amount of different folding routes and it also explains the exponential folding kinetics observed within experiments.

However they predicted also a Type 0 or downhill scenario of protein folding, in which the conformational entropy compensates the stabilization energy along the reaction coordinate in such a way that just one minima exist close to the native state. Here the protein might fold without any or with a marginal small barrier. In this case a 'bottleneck' wont exist and therefore multiple exponential folding kinetics would be expected.

1.5 Molecular Dynamics Simulations of protein folding

In parallel to the lattice [11, 12] or Go [26] models to describe the protein folding process computationally, many effort has been made to conduct atomic level molecular dynamic simulations in equilibrium [27] beginning with the pioneering work of Levitt, Warshel [28] and Karplus [29] in the mid 1970's. Here both the protein and the solvent are represented at atomistic detail, with energy components describing variations in energy due to bond

and angle stretching, torsional rotations and long range electrostatics. All these parameters are implemented inside a force field, which is used to conduct the folding simulation of a protein in equilibrium by solving newtons second law for each atom in the system (protein and solvent) to dictate its trajectory at each timestep, which is usually in the range of 1-2fs. Several different force fields have been developed over time like the CHARMM or the AMBER force field [30]. For a long period these all-atom protein folding simulations lacked of computational power to conduct simulations in a reasonable amount of time to finally reach (un)folding times in the μs-ms range of fast folding proteins. The MD simulation time reached was for a long time in the range of nanoseconds, therefore still far away from the real protein folding time.

Here a important breakthrough was reached by the E.Shaw group and the development of a specialized supercomputer, called Anton, which was tremendously increasing the computational power of MD simulations and therefore decreasing the overall calculation time [31, 32]. In 2011 this machine than was proved to calculate the (un)folding trajectories of nine fast folding proteins and three more stable protein using a modified CHARMM force field [33]. This machine enabled to calculate (un)folding trajectories in the ms range (up to 10^6 times faster than before) and many transitions from folded to unfolded state could be observed for the twelve chosen protein structures. From this point MD simulations showing the whole folding trajectory of a fast folding protein are more and more used to compare with theoretical models [34] and with experimental results [35, 27, 36].

Thus these latest advances in conducting computational calculations make the all-atom computer simulations a promising tool to complement protein folding experiments (which will be discussed in the next section), especially with using the calculation capacity of graphics processing units GPUs adapted from the video gaming industry and further developed force fields [37, 33, 38]. Therefore together with the theoretical and experimental description of protein folding, MD simulations could be able someday to be a reliable tool for prediction protein folding from an unknown amino-acid sequence. Therefore still many effort has to be done to tune the used force field in order to predict all of the experimentally observed protein (un)folding behavior.

1.6 Protein folding experiments

1.6.1 Bulk experiments

Much effort has been done through the last decades in order to test experimentally the discussed protein folding landscape predictions and the results from simulations [3]. Spectroscopic methods like fluorescence, absorbance or circular dichroism enabled to monitor the folding/unfolding reaction of the protein sample in solution (in bulk, i.e. multiple molecules) depending on the amount of denaturant used (e.g chemical denaturant or temperature). In these equilibrium measurements, the change in the fluorescence signal of the protein upon chemical denaturant titration results in a sigmoidal curve. From there the free energy of folding in water ΔG_{H_2O} can be estimated [39].

Another important technique which is widely used to investigate protein unfolding is differential scanning calorimetry (DSC). DSC thermograms (heat capacity vs temperature) of a protein solution will give an idea about the energy storing capacity of the protein during unfolding. Therefore in the temperature range where protein unfolding occurs, the heat capacity is high and the protein unfolding process is indicated by a peak in the DSC thermogram. From the DSC thermogram, the unfolding caliometric enthalpy (calculated from the area under the heat capacity peak) and the van't Hoff enthalpy (calculated from the shape of the DSC peak) is estimated [39].

Kinetic measurements allow the estimation of the folding (k_f) and unfolding rate (k_u) of the protein at a presence of the denaturant concentration (m). Here, the fluorescence signal is monitored after a rapid perturbation, which forces the system to establish a new equilibrium point. From a so-called chevron plot ($\log(k_f, k_u)$ vs m) the (un)folding rates at zero concentration can be estimated, which allows also the determination of free energy of folding in water ΔG_{H_2O} [39]. However this approach of kinetic measurements requires always a kinetic model in order to interpret the data.

To avoid the restriction that a kinetic model is needed to interpret the data, the thermodynamic barrier height can be estimated directly from DSC measurements. This approach consists of analyzing the heat capacity data with a phenomenological 'Variable-barrier model' introduced by Munoz and Sanchez-Raiz, which is based on the classical Landau theory for critical transitions [40, 41]. It is based on the argument that for proteins with low energy barriers, conformations at and near the the top of the free-energy barrier also contribute to the equilibrium properties of the system. I.e protein folding is not described as a sequence of macroscopic states (i.e unfolded and native state) but more as a continuous process, which involves an ensemble of microstates. Thus the advantage of the

'Variable-barrier model' is that the height of the free-energy barrier and general properties of the folding ensemble are not preassumed but rather directly obtained from the experimental data. This method has also been compared to the free energy barrier heights estimated from kinetic data, showing good agreement [42].

The later development of fast protein folding experimental techniques (LASER induced temperature-jump techniques [43] and NMR relaxation spectroscopy [44, 45]) increased the resolution time of kinetic measurements to picoseconds. That enabled to measure the speed of the formation of structural elements of proteins from simple peptide bond rotations (1-2ns;[46]), α helices (200ns;[47]) to β hairpins (1-5 μs;[48]). Additionally hydrophobic chain collapse was estimated to occur within 100ns [49]. Figure 1.7 gives an overview about the timescales of various stages of protein folding from α helices formation to the complete folding of ultrafast folding proteins [50].

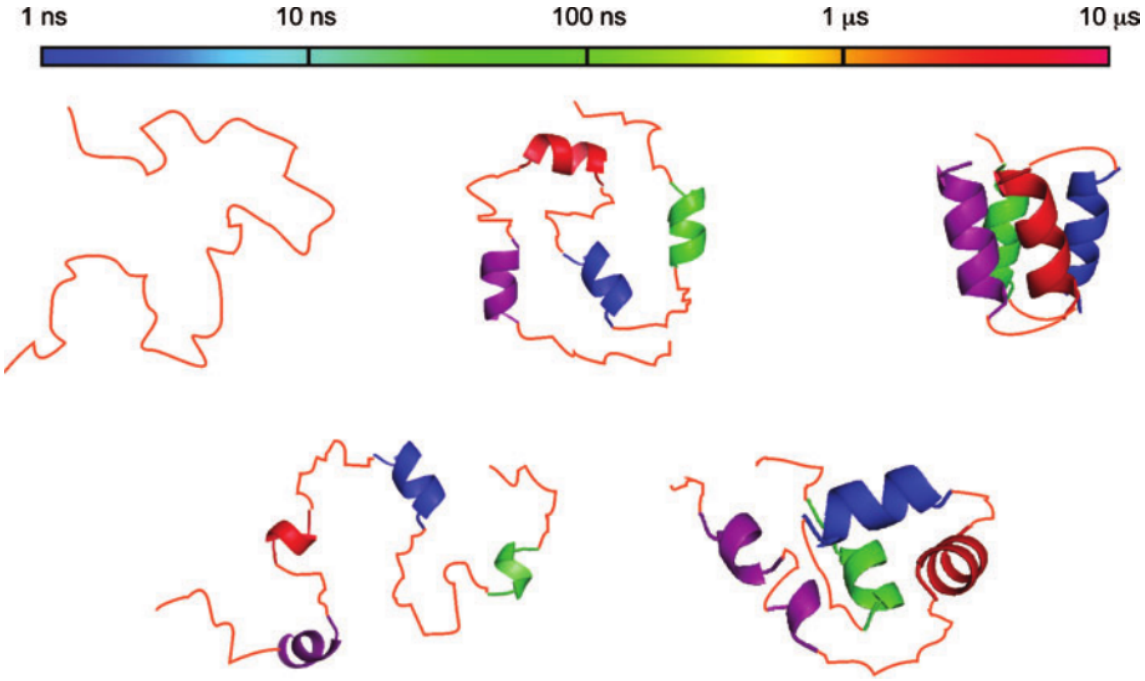

Figure 1.7: Timescale of various stages during the protein folding process of ultrafast folding proteins. Figure taken from [50]

.

The ability to study the folding mechanism of the fastest folding proteins (few miliseconds and faster folding times) [50, 51, 52] has important practical implications. Because their folding barriers are sufficiently small (<16RT) it becomes possible to resolve experimentally folding conformational ensembles and their intrinsic dynamics.

These experiments raised the question of how fast a protein could possible fold, i.e. is there a speed limit for protein folding? Kubelka et al [51] concluded that the generic

speed limit is about $\frac{N}{100\mu s}$, where N is the number of residues. Therein the speed limit is set by backbone diffusion and chain length, which enables small globular proteins to fold within a 0.5 - 5μs timescale. Another approach to estimate the speed limit from the kinetic data is to use an approximate form of the pre-exponential factor in Kramers theory (see Kramers equation 1.2) [51, 53].

Both the founding of a speed limit for protein folding and the development of nanosecond kinetic techniques increased the search for ultrafast proteins [54] and also the engineering of proteins to increase their folding speed [55]. In this context protein properties were searched for, which could drive the protein folding rate to higher values. Therein protein topology (relative contact order) has been found to effect the folding rate most [23]. The effect of protein size [56] and protein structure [57] was shown to have also effect on the protein (un)folding rate and stability. Therein a 1D free energy surface model using the nativeness n as a single reaction coordinate developed in the Muñoz laboratory [58] was able to predict the protein kinetic values. One-dimensional projections have also been successfully applied to the prediction of two-state folding rates from protein structures [59] and more recently to reproduce even the complex folding process of a multidomain protein [60].

1.6.2 Single Molecule Measurements

The main advantage of using single molecule techniques is that it allows the detailed examination of heterogeneous populations without ensemble averaging (when investigating protein folding in bulk). In the late 1990's recent technological advances made it possible to investigate protein folding on a single molecular level [61, 62]. Particularly single molecule fluorescence (e.g. FRET [63, 64]) and single molecule force spectroscopy (AFM, magnetic and optical tweezers [65]) enabled the observation and tracking of the dynamics of single proteins. Because single molecule force spectroscopy will be introduced in detail in the next chapter, here in this section the focus will be on single molecule fluorescence spectroscopy. Förster/fluorescence resonance energy transfer (FRET) describes the non-radiative transfer of singlet excitation energy from donor to acceptor dye molecules. This energy transfer is strongly distance dependent (30-70 Angstrom), which makes it very attractive strategy to investigate conformational changes within protein dynamics. Herein the protein of investigation will be labeled with a donor and an acceptor dye in a suitable manner. For example in order to study protein folding, when the protein is folded there will be a detectable signal (short distance between the dyes) whereas when the protein is unfolded the

signal will be lost (long distance between the two dyes). The probability that a photon absorbed by the donor will lead to energy transfer to the acceptor is called the FRET efficiency E.

There are two methods for to perform FRET experiment for studying protein folding. One method is to carry out experiments on freely diffusing molecules through the confocal volume and the other method is to immobilize the protein on a surface [66] or to encapsulate them in lipid vesicles [67]. Whereas the freely diffusing method can reveal conformational subpopulations within the protein sample, the immobilization method offers the investigation of folding/unfolding trajectories of one single protein. In the later case the kinetics can be directly estimated from the dwell time distribution of the individual states of the protein.

Recent experimental setups enabled to resolve the transition state ensemble during (un)folding of ultrafast proteins [68, 69]. Further information about FRET can be found elsewhere [70].

1.6.3 Experimental identification of two state and downhill folding

The experimental identification of ultrafast folding proteins with folding rates in the range of $(10^6 - 10^4 s^{-1})$ emphasized that proteins folding at the speed limit could exert downhill folding processes [52]. The consequence for proteins folding at the speed limit is that some might fold downhill (One-state folding; Type 0) without having an folding barrier (two-state folding; Type I) [13]. However just by the identification of proteins folding close to their estimated speed limit does not necessarily mean that they fold downhill e.g. as in the case of Vilin headpice subdomain, where a small energy barrier still persists (VHP) [53].

Two state folding; Type I

Many small proteins investigated experimentally appear to follow a fast, two-state folding reaction [71, 72]. That means the barrier to folding is low (2-8 $\frac{kcal}{mol^{-1}}$ up to 16 RT ($k_B \mathrm{T} \cdot N_A$) and the populations of the protein molecules at any stage of the reaction (e.g. amount of denaturant or temperature) are either in the denatured or fully folded state. The reaction is described as a simple U $\rightleftharpoons$ F pathway, with a barrier height (ΔG_{U-F}) and a folding (k_f) and unfolding rate (k_u).

Single molecule FRET experiments revealed also many fast folding proteins to fold two-state, like the cold shock protein TmCsp [64]. Here the two-state behavior is stated in

detecting two peaks in the transfer FRET efficiency histograms, which corresponds to the unfolded and folded subpopulation [73] and in detecting just two states within the time trajectory of a single TmCsp protein [67].
However it seems that even the end states of some two-state folding proteins are not structurally invariant upon chemical denaturation, which was e.g revealed for clear two-state proteins like SH3 from α-spectrin. Here an expansion of the conformational ensemble was shown in the native and unfolded state in using single molecule FRET and NMR spectroscopy[74].

One State Folding, Type 0

The energy landscape theory predicts the case in which proteins can fold downhill (one state folding). Here the energy barrier has been removed either by natural selection or by protein engineering, leaving fluctuations below 2-3RT ($RT = k_bT \cdot N_A$). The advantage of finding or constructing proteins with such a downhill folding scenario is that it can give the possibility to observe all intermediate states progressing from the unfolded to the native state [75].
Hence the discovery of downhill folding proteins would provide the unique opportunity to resolve in detail structural and kinetic properties of the conformational ensemble with respect to the amount of denatural stress used, which cannot be identified in the two-state scenario. A probe dependent behavior in equilibrium unfolding experiments might be the best indicator of a protein folding mechanism over a marginal free energy barrier [22]. Because in the downhill scenario the conformational structure changes gradually, different probes should measure different signals during unfolding. The resulting unfolding signals from different spectroscopic techniques should be still sigmoidal but in contrast to two-state unfolding, the transition is much broader and the measured midpoint at T_m wont be identical for different probes used. E.g. different probes switch from the native to the unfolded signal at different positions along the reaction coordinate(s).
Downhill folding proteins have been experimentally identified using a combination of spectroscopic techniques (both equilibrium and kinetic measurements) and calorimetry experiments. Here BBL (the peripheral subunit binding domain (PSBD) from Escherichia coli's 2-oxogluterate dehydrogenase multienzyme complex) was identified to fold globally downhill (one-state) [76, 77] and for example the gpW protein domain from bacteriophage λ was found to fold in a nearly downhill fashion [78]. The findings for the gpW protein that it has a marginal small energetic barrier and properties of a nearly downhill folding protein

were further strengthened through a NMR relaxation experiment, revealing a complex free energy landscape [79].

Further work identifying characteristics of a global one-state downhill folding protein was done by Gruebele and coworkers [55, 80]. Therein it was shown experimentally that two kinetic phases exist in a set of mutants, which where increasing the folding rate inside the five-helix bundle protein λ_{6-85} and therefore reaching the speed limit. Here also the probe dependence of downhill folding proteins was confirmed later [80].

Leeson et al [81] observed a similar diffusive, downhill folding process for the small β-sheet cold shock protein CspA from *Escherichia coli* in using T-jump IR spectroscopy. Here the protein normally was shown to fold at the transition midpoint with single exponential kinetics stating a clear 2-state process. However under strongly folding conditions (larger T-jumps, >12°C) a fast, temperature independent and highly diffusive process was observed with high nonexponential kinetics. Here single exponential and biexponential kinetics were described to coexist within one protein sample.

Another experiment indicating continuous and probe dependent unfolding was detected by measuring FRET signal in bulk and the kinetics of the small monellin protein [82].

A further experimental milestone was that the entire folding process of downhill proteins was thought to be resolvable with high resolution equilibrium experiments with NMR, because protein population can be high anywhere along the reaction coordinate. Sadqi et al. [83] exploited that idea in analyzing the equilibrium thermal unfolding process of the BBL protein. Monitoring the chemical shifts as a function of temperature for 158 protons of the BBL revealed high structural heterogeneity during downhill folding. Similar results were obtained by the same atom-by-atom analysis of the above mentioned nearly downhill folding gpW protein [35].

The before mentioned 1D free energy surface model [58] was used to reproduce and quantify the systematic deviation from the two-state behavior of temperature and chemical kinetic measurements of ultrafast proteins. As a result proteins crossing marginal barriers (<3.5RT or <9 kJ mol^{-1}) should fold in a downhill fashion in native like conditions, while proteins crossing barriers higher than 3.5RT should fold in a two-state fashion.

FRET single molecule experiments with BBL revealed a single conformational ensemble at all denaturant conditions used. Furthermore it highlights that the downhill folding scenario is not necessarily associated with ultrafast kinetics [84]. In order to conduct these experiments the folding speed of BBL was needed to be slowed down $\approx$ 200-times by a combination of chemical denaturant and low temperature, in order that the folding/unfolding

rate of BBL is above the binning time of the Instrument. A special photoprotection cocktail [85] was additionally needed in order to conduct freely diffusive FRET experiments with a 50μs resolution (binning time).

Thus experimental results indicate the existence of proteins, which fold in a downhill (one-state) or nearly downhill (marginal barrier) manner.

Chapter 2

Mechanical unfolding of proteins

This following chapter introduces the wide field of single molecule force spectroscopy (SMFS), being part of the single molecule techniques to investigate protein dynamics as discussed before. Terms and definitions will be given in detail about the usage of the atomic force microscope (AFM) in the field of SMFS, as it is the technique of choice in this thesis.

2.1 Mechanical forces in biology

Mechanical force play a fundamental role in biological processes [86]. These mechanical processes are involved in nearly every facet in the cell cycle, ranging from cell mobility processes, molecular motors to replication and segregation of the DNA. Within this aspect a wide range of examples has been found in which a mechanical force is directly applied to a protein, i.e. proteins exist which can resist and also provoke mechanical forces in the pN range.

For example the largest protein found in the human body (4200kDa) is the protein titin consisting of many single domains, being responsible for passive elasticity in the skeletal and cardiac muscle sarcomere. Therefore it was one of the primary targets to conduct protein force spectroscopic studies [87]. Many so-called molecule machines have been identified, which are built-up of complex assemblies of many proteins and convert chemical energy gained from ATP into real mechanical work. The recently experimentally described molecular motor CLpXP has been found to apply a mechanical force in order to denaturate proteins inside the cell [88].

Whereas direct measurements of force in vivo inside the cell awaits its technical development, much can be gained by studying the mechanical response of a single protein in vitro.

Here the ability to study mechanical processes in real time but especially the ability to apply mechanical forces in the range of pN to a single protein and to detect its reaction, has been revolutionized by the development of new single molecule experimental techniques over the last 30 years. Hence the energy landscape of a protein can be investigated in using mechanical stress, a biologically important perturbation.

2.2 Single Molecule Force Spectroscopy Experiments

Basically three techniques have been evolved so far and are commonly used in the new field also called bio-nanomechanics, as this techniques allow the detection of forces in pN and displacements in nm. This is normally the range at which for example molecular motors work and proteins get mechanically unfolded [88]. These techniques are the optical and magnetic tweezers and the atomic force microscope[65].

2.2.1 Optical and Magnetic Tweezers

Optical Tweezers (OT)

The fundamental principle of the optical tweezers is that electromagnetic radiation exerts a pressure on any surface, defined as radiation pressure. Furthermore when the gradient optical forces exceed those from scattering, an object can be stably trapped. Based on that idea in 1986 Ashkin built the first optical trap, caused by focusing the light from an Argon LASER (λ=514.5nm) with a high numerical aperture (NA=1.25) lens, which exerts forces in the order of picoNewtons on dielectric nano-particles used (μm-nm range) in a fluid chamber [89]. Further instrumental development enabled the detection of displacements in the nanometer range, thus enabling the tracking of biological molecular processes [90, 91]. In a typical experimental setup one side of the molecular probe (e.g. DNA/RNA; protein; molecular motor) is attached on a micron-sized dielectric bead particle, which is than captured in the optical trap, while the other side of the probe is immobilized, either on a fixed surface or again on a micron-sized bead particle, which is then hold either by a micro-pipette or by another optical trap. A mechanical force on the molecular system is now applied by moving one fixed bead particle.
This setup enables different modes of applying a mechanical force on the molecular probe. Experimental options are either to move the bead particle with a constant velocity or to held the mechanical force applied on the bead particle constant [92].
Typically, optical tweezers enable the application of mechanical forces in the low pN-range

(0.1-100pN) and they allow a mechanical displacement accuracy in the sub nanometer range (0.1-10^5nm), through the usage of sensitive position detectors, based on quadrant photodiodes. The time resolution which achieved so far is in the sub milisecond range [93].

Magnetic Tweezers

The idea of using magnetic forces as a single molecule technique was developed in parallel to the optical tweezers [94]. Magnetic tweezers are based on the principle that a magnetic bead with a magnetic dipole moment experiences a force when being inside a magnetic field of a pair of a permanent or electromagnet.
The mechanical forces exerted by magnetic tweezers can be very small and can reach the femtonewton range (0.0001-100pN). Magnetic tweezers experiments are just capable in operating in the constant force mode. Additionally to the application of a mechanical force in one dimension, magnetic tweezers can apply a torque. Like for the optical tweezers the molecular probe is attached between the magnetic bead and a glass surface through specific binding. The length and time resolution is typically in the several tens of nanometer (displacement) and milisecond (time) range.

2.2.2 Atomic Force Microscope

The atomic force microscope (AFM) [95] is a derived version from the scanning tunneling microscope (STM), which was invented by Binnig and coworkers in 1982 and was meant for imaging non-conductive surfaces [96].
However quickly after the AFM was invented, it became one microscopic technique of choice in the field of Biophysics/Biology, when imaging or manipulating the surface of biological probes like cells, DNAs or single proteins adsorbed on a surface [97, 98, 99, 100, 101]. This is because the AFM can operate in a non-adhesive mode and under quasi physiological conditions with a sub-nm image resolution capacity.
The principle of AFM imaging is to probe the surface with an ultrasharp tip (tip radius: tens of nanometers, material: silicon nitride) mounted on a cantilever. Either the cantilever or the surface is being moved by an electrical PIEZO acutator in a sub-nm range in x-y-z directions. When moving the tip in contact over the surface (contact mode), the tip will bend and the signal on a photodiode caused by a diode LASER beam reflected on the back of the cantilever will change depending on the bending on the tip.
The change of the signal on the photodiode is then directly converted into a topographical height image. There exist also a non-contact mode, in which the tip is stimulated to

vibrate with a given frequence and amplitude setpoint (Tapping Mode). When now the tip comes close to the surface it will experience atomic forces increasing or decreasing the given setpoint amplitude, detected on the phododiode. Using a feedback-loop to maintain the given amplitude setpoint will now cause the PIEZO to move in z-direction, which is translated directly into a topografical height image. The Tapping Mode has the advantage that it does not harm or destroy the biological surface. Figure 2.1 shows the AFM setup schematically.

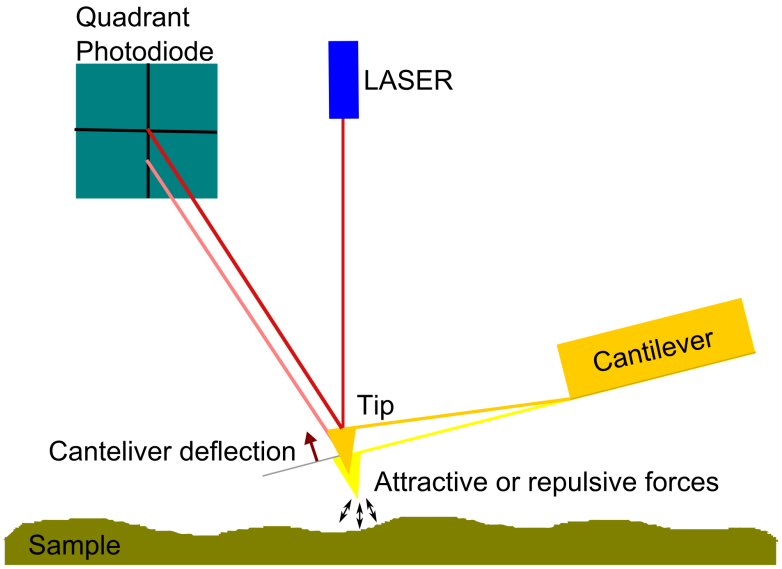

Figure 2.1: AFM setup: A quadrant photodiode receives a signal from the reflected LASER beam on the backside of the AFM cantilever. Attractive and repulsive forces near the surface of the sample cause the AFM cantilever to bend in the non-contact mode. This deflection then causes a change in signal of the photodiode. Either the sample or the cantilever is being moved by a PIEZO acutator.

.

The resolution of the AFM is mainly restricted on the one hand to mechanical limitations like the PIEZO movement and the sharpness of the tip and on the other hand to electronical processing response times. Recent advances in of AFM imaging like producing ultra sharp tips and electronic processing elements with faster response times made it possible to track e.g. the movement of myosin V on actin with a special developed high speed AFM [102, 103].

Single Molecule Force Spectroscopy using the AFM

The idea of using the AFM setup to measure and apply mechanical forces of biological samples started with the investigation of forces between complementary strands of DNA in 1994 [104]. In 1997 then the AFM setup was used for the first time to apply a mechanical

force on the protein sample titin (Immunoglobulin Domains) [105]. For single molecule force spectroscopy measurements the cantilever will be moved in z-direction towards the surface on which the protein sample is adsorbed. After touching the surface and retracting the cantilever, a mechanical force will be applied on the protein, if the protein indeed was adsorbed on the tip. This force resistance will cause the cantilever to bend until the protein is being unfolded (broken) or detached from the tip and will release the bending of the cantilever.

However because the cantilever tip can also cause unspecific bending signals when being in contact with the surface usually a polyprotein construct is used, which has to be built with molecular biology techniques. This polyprotein construct consist of several 6-12 domains of the protein of investigation. Applying a mechanical force on the polyprotein construct will then result in well a distinguishable repetitive pattern, therefore helping to identify the signal from the protein of investigation.

Mostly non-specific binding of the protein sample to the substrate is used for force spectroscopic measurements using the AFM. One strategy is to equip the polyprotein construct with one or two single cysteine residues in the end, as it is known that cysteine residues form a disulfide bond with gold, which is often used as a substrate. On the other side the polyprotein construct binds unspecifically to the cantilever through a process which is not well understood. However it has been suggested that the unspecific binding to the cantilever is caused by the high pressure exerted by the sharp tip on an area in the nm range. Instead of a gold surface also a Ni-treated glass surface can be used, where the covalently bonded Ni binds to the HisTag-tail from the polyprotein construct, which is normally used for the purification of the polyprotein construct [106]. However also specific binding strategies to both the substrate and the cantilever exist[107]. Herein, especially the so-called HaloTag thethers were described to work promising within the AFM setup, allowing to repeat mechanical force cycles on one and the same polyprotein construct for up to hours [108].

Thus in contrast to force spectroscopy measurements done with optical or magnetic tweezers, first a polyprotein construct has to be used as a sample in order to have a clear and distinguishable pattern and second unspecific binding strategies can be applied to attach the sample to the cantilever.

Force calibration

Before starting a force spectroscopy experiment (AFM, optical or magnetic tweezers) the force needs to be calibrated in order to transform the measured electrical signal on the photodiode (in volts) into a force (in newtons). In the case of the AFM the spring constant of the cantilever needs to be estimated. There are several ways of calibrating the force from the measured deflection of the cantilever.

For force calibration the cantilever is seen as an hookean spring. The most convenient way of estimating its stiffness k is to measure its free oscillations far from the surface caused by the thermal fluctuations. For small bending angles the system can be seen as an harmonic oscillator, which will fluctuate in response to thermal noise. By knowing the Hamiltonian of the system and applying the equipartition theorem 2.1 the spring constant k can be estimated [109]. The equipartition theorem states that the kinetic energy of each degree of freedom (such as a vibrational mode) equals half the thermal energy k_BT

$$\langle \frac{1}{2} m\omega_0^2 q^2 \rangle = \frac{1}{2} k_B T \tag{2.1}$$

Here $\frac{1}{2}m\omega_0^2 q^2$ is the kinetic therm of the Hamiltonian, where q is the displacement of the spring, m the oscillating mass and ω_0 the resonant angular frequency of the system. With k=$m\omega_0^2$, the spring constant can then be obtained by measuring the mean-square displacement of the harmonic oscillator k = $\frac{k_BT}{\langle q^2 \rangle}$. Practically for estimating the spring constant the thermal fluctuation of the cantilever will be measured and fourier transformed to gain the the Power Density Spectrum of the cantilever. As the integral below the power density spectrum equals the mean square displacement of the fluctuations $\langle q^2 \rangle$ and is being measured in units of V^2, the spring constant can be estimated.

Additionally in order to calibrate the force, the measured signal from the photodiode in volt has to be correlated to the bending (displacement) of the cantilever in nm. Therefore the slope of deflection vs extension trace in the contact region of the cantilever to the surface will be measured, which correlates the change in voltage of the PD to the change in nm of the displacement of the cantilever. To estimate this slope with a high accuracy stiff surfaces need to be used like gold or glass.

Therefore finally the spring constant k can be estimated by k=$\frac{k_BT}{\langle q^2 \rangle s^2}$ With knowing the spring constant k the force F can be estimated by applying the Hookean law, wherein the the force is proportional to the extension.

Constant velocity measurements

In the constant velocity mode, the cantilever is being approached and retracted from the surface with a constant velocity, thus applying an uncontrolled mechanical force on the protein sample. The first AFM force spectroscopy measurements were done in this mode [105]. Using a polyprotein construct as a probe will result in a so-called sawtooth pattern in the corresponding force-extension trace (Figure 2.2), in which each peak corresponds to the unfolding of a single domain inside the polyprotein construct.

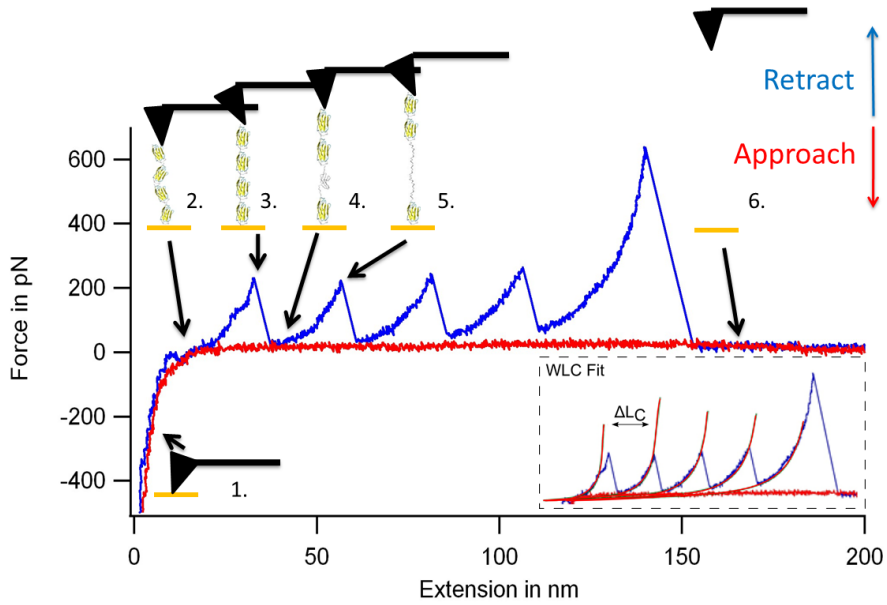

Figure 2.2: A force extension curve is shown for the case of the titin I27 domain. Typically the approach trace is shown in red and the retract in blue. Here the retract trace shows four unfolding peaks of the titin I27 domain, the last peak corresponds to the disruption of the protein sample and the cantilever tip. The sketches numbered from 1 to 6 indicate the corresponding position of the cantilever with respect to the surface and the state of unfolding of the polyprotein sample. The insert shows the fit of each unfolding peak to the WLC-model.

The figure shows the example of the unfolding of four titin I27 domains [87]. The sketches in the figure numbered from 1 to 6 indicate the corresponding position of the cantilever to the surface and the state of unfolding of the polyprotein sample. First the cantilever approaches the surfaces, where the polyprotein sample is adsorbed and gets bend when in contact (1). In the positive event that the cantilever picks up a polyprotein chain, the chain will be stretched while the cantilever is retracted from the surface (2) until a applied force is reached close to the unfolding force of the protein domain (3). This raise in force is results the first peak in the corresponding force vs. extension curve (see figure 2.2). At this maximum force one of the domains inside the polyprotein will mechanically unfold

and therefore release it containing aminoacid chain. This will cause the cantilever to bend in the opposite direction and therefore the force will drop in the force extension trace (4). Now while further moving the cantilever apart from the surface at a constant velocity the polyprotein chain will be stretched again until again the unfolding force is reached (5), which results in the second peak. This will be repeated until the whole polyprotein chain gets detached from the cantilever (6), which results in the last force peak, which often is much higher than the single unfolding peaks of the protein sample. The rate of picking up a single polyprotein chain depends mostly on the concentration of the sample and on the technique used to attach the sample on the surface and on the cantilever, which will be explained in the section about binding strategies. As it can be seen in the trace in figure 2.2 the force will not drop to zero after the unfolding of a protein domain, which become more clear the more domains have been unfolded before. This is because of the entropic force of the already unfolded polyprotein chain which avoids the cantilever from reaching the zero deflection position.

For analysis each unfolding peak of the sawtooth pattern is then being fitted to the so-called Worm Like Chain (WLC) equation [110] (see insert in figure 2.2)

$$F(x) = \frac{k_B T}{p}\left[\frac{1}{4}\left(1-\frac{x}{L_C}\right)^2 - \frac{1}{4} + \frac{x}{L_C}\right] \tag{2.2}$$

With p being the persistence length, k_B the Boltzmann constant and T the temperature. This equation describes the entropic elastic properties of a polymer under force with a specific contour length L_C and it has been found to fit well also a protein under force as a first approximation [105] and also when stretching DNA [111, 112], using a persistence p of 0.4nm which is approximately the length of one amino-acid residue. Another model frequently used to analyze the constant velocity data is the freely jointed chain (FJC) model [113], however usually the WLC model fits well as a first approximation. The difference of the estimated contour length of two consecutive unfolding peaks (ΔL_C) corresponds then to the released relative amount of aminoacids during unfolding and is therefore a typical fingerprint of the protein sample used beside its unfolding force.

Force clamp measurements (constant force and force ramp)

However also like in the optical tweezer setup, the applied force on the protein sample can be controlled with a feedback loop. This so-called force clamp AFM setup was developed in the lab of Julio Fernandez [114, 115]. With the force clamp AFM the deflection of the cantilever and therefore its exerted mechanical force on the protein sample can be hold

with time at any given setpoint, furthermore the force can also be changed linearly with time (called a force ramp) [114, 116]. Basically any force sequence can be programmed. Within these measurements, using a polyprotein construct will result in a typical staircase pattern in the corresponding length- and force vs time trace (see figure 2.3).

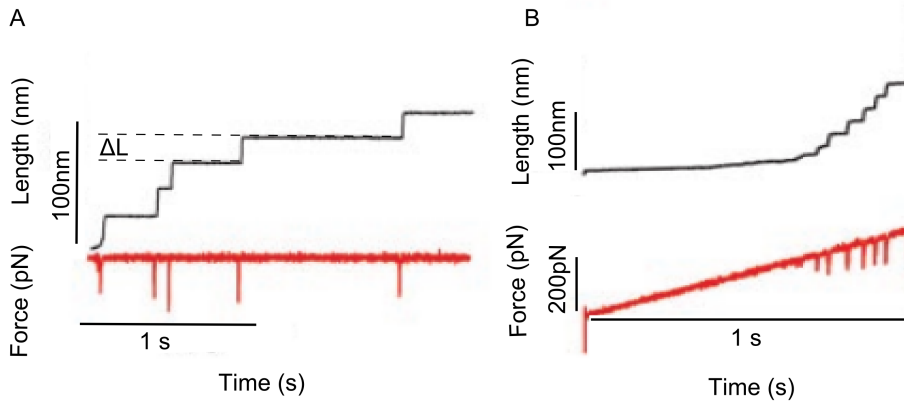

Figure 2.3: (A) A typical constant force trace of a polyubiquitin construct sample is shown at a constant force of 120pN. After applying 120pN for a certain amount of time one of the ubiquitin domains will mechanically unfold by chance and cause the PIEZO to move a step length in order to hold the given setpoint of 120pN. The step length ΔL in the length vs time trace therefore corresponds to the unfolding length of the ubiquitin domain.(B) A typical force ramp trace is shown in which the polyubiquitin construct is set under linear increasing force at a rate of 300pN/s. Here by chance the ubiquitin domains unfold stochastically one after another, when the force is in the range of the mean unfolding force of ubiquitin. Both traces were taken from [116].

.

2.3 Mechanical protein (un)folding experiments

The atomic force microscope is widely used to investigate protein (un)folding under a mechanical force [117]. Here a high variety of proteins have been studied already, which enabled to classify proteins into mechanical resistant and non-resistant proteins and also to identify mechanical resistant structural motifs. Proteins have been found to unfold within a few 10-20pN piconewtons (non mechanical resistant) e.g. like ankyrin [118], spectrin [119], β-catenin [120] or talin [121] up to several hundreds of pN like titin [122], ubiquitin [116], Cohesin I [123], GFP [124], pili [125] or tenascin [126]. The upper limit of mechanical resistance can be rigorously thought to be the force at which a covalent bond breaks, which is estimated to happen in the nanonewton range [127]. Beside the classification into mechanical resistant and non resistant proteins, it has been found that protein without any known physiological mechanical function can be highly mechanical

resistant like the cold shock protein Csp [128] or the protein L [129].

So far protein (un)folding studies revealed a broad distribution of mechanical stabilities (see e.g database of experimental AFM constant velocity data [130]). Structural motifs like the alpha helix has been found to have less mechanical resistance than proteins made of beta strands. Especially parallel β-strands (also called mechanical clamp motif) have been identified to be able to bear the high mechanical forces like in the case of titin, ubiquitin, protein L, pili or Cohsin I [131]. Most proteins studied so far unfold and fold without showing any intermediate with some exeptions like ubiquitin [116], fibronectin [132] ,T4 lyzosyme [133] and also in the titin I27 domain [134].

Optical and magnetic tweezers so far are mostly used to study the RNA, DNA stability under a mechanical force. This enabled fascinating investigations on identifying and tracking molecular machines acting in real time on DNA/RNA [135, 136, 137, 138, 139]. Fewer experiments concentrated on the study of the mechanical response of a single protein [140, 141].

Recently there is a change in using the optical and magnetic tweezers for studying protein (un)folding of less resistant proteins in order to detect small conformational changes in the low pN range (<10pN) [142, 143, 144, 145, 93]. This is basically because the force control and resolution enables the application of mechanical forces as low as 0.01pN (magnetic tweezers) or 0.1pN (optical tweezers) which is of up to three magnitudes lower than what can be achieved in the AFM measurements, which lowest applicable force is around 10pN. For optical tweezers this is because optical traps and AFM cantilevers have very different spring constants, in the order of 0.1 pN/nm vs 10pN/nm respectively. Softer springs have better force resolution and can access lower force regimes.

Beside the unfolding of a protein also the folding (refolding after unfolding) of a protein can be investigated during AFM force spectroscopy measurements, like it is frequently done for optical tweezers experiments with proteins [146]. During both constant velocity [126, 118] and constant force measurements the applied force can be reduced and refolding can be observed[147]. During constant velocity measurements at a extremely low velocity of 1nm/s using a low-drift AFM and a cantilever with a small spring constant of 6pN/nm, Matthias Rief et al [148, 149] could observe folding/unfolding transitions and binding/unbinding events of calmodulin. During constant force measurements hopping of a protein is so far just observed during optical tweezers experiments [141, 142, 143, 144] due to the high spring constant of the AFM cantilever compared to the spring constant of an optical trap, however Julio Fernandez et al [150, 151] were investigating in detail

the refolding of ubiquitin and the titin I27 domain into a collapsed state until they finally refold. Here within force clamp measurements the force on the protein sample can be easily decreased due to the feedback loop. This enables the protein sample to refold, if the release of force and the amount of time at the released force is sufficient enough.
In sum, both types of measurements in constant velocity mode (far from equilibrium) and in constant force mode (within equilibrium) allow to reveal kinetic information like (un)folding rates k_U, k_F and the change in free energy of the protein of investigation. These parameters than describe the mechanical stability of the protein under force and its underlying energy landscape at zero force can be derived as it will be shown in the next two sections [152].

2.4 The effect of force on the free energy landscape

A mechanical force applied on a protein changes its thermodynamic stability and can speed up or slow down its kinetic rates [153, 86]. Figure 2.4 depicts the view of that an applied force tilts the free energy surface G down to the unfolded state with the end-to-end distance x (of the protein) as a reaction coordinate.

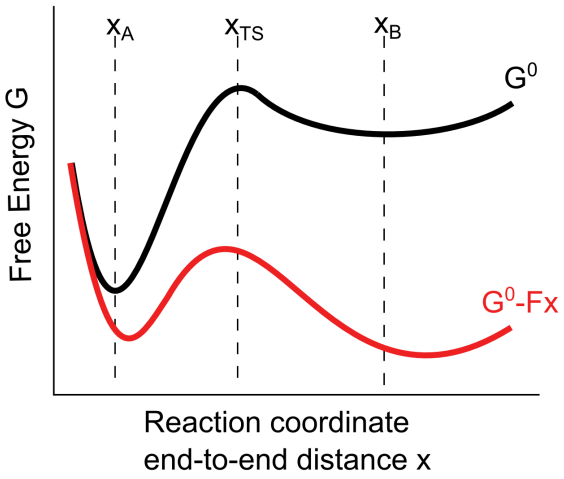

Figure 2.4: Sketch of how the free energy surface G (black) is tilted by the application of a mechanical force in the direction of x (red). Beside the position of the state A,B and TS also the curvature of the harmonic wells of the state A,B and transition state TS will be affected

.

As the force increases the probability for the protein to populate the unfolded extended state also increases. Thus the barrier between folded and unfolded state is lowered and also the height of the unfolded state itself will be lowered to a value below the folded state. As it will be discussed in the next chapter, also the curvature of the harmonic wells in

state A and B and for the transition state TS are usually affected by the application of a mechanical force.

The effect of force on protein kinetics

The effect of force on the kinetics of a protein, was first investigated by Bell [154]. He predicted that protein-protein bonds would rupture at a rate that depends exponentially with the force. As a result, the forward and reverse rates described by Kramers equation (see 1.2) are tilted or modified by the external force F in direction x (pulling direction). For a two-state system the dependence of the folding and unfolding rates (k_f, k_u) are then described in the following Bell equation.

$$k_f(F) = k_f^0 e^{-\frac{F\Delta x_f}{k_B T}}, k_u(F) = k_u^0 e^{\frac{F\Delta x_u}{k_B T}} \tag{2.3}$$

Where $k_{f/u}^0$ is the folding/unfolding rate constant in the absence of force, F is the stretching force, $\Delta x_{f/u}$ is the distance to the folded or unfolded transition state, k_B is the Boltzmann constant and T is the absolute temperature. The Bell formulation can be used as a first approximation for deriving the (un)folding rate of the protein at zero force and the distance to the (un)folded transition state.

Reconstructing folding energy landscapes by single molecule force spectroscopy

Depending on the type of force spectroscopic measurements on a protein (constant force at near equilibrium or constant velocity at non-equilibrium), the strategies from extracting thermodynamic and kinetic properties to finally describe its 1D energy landscape differ [155]. Furthermore they can be also divided in model-free and model-dependent approaches.

Therein the most convenient non-model approach is to reconstruct the energy landscape directly from the probability distribution P(x) of the molecular extension x at the equilibrium force F$\approx F_{1/2}$, revealed from equilibrium constant force measurements, which can be easier accessed by optical tweezers than by AFM measurements because of the much lower spring constant [156]. Here $F_{1/2}$ is the mechanical force at which the protein populates half of its time the extended unfolded state and the other half in the folded state.

However in a case of a protein with a slow unfolding rate or having a too stiff spring constant (AFM), the constant force approach with F$\approx F_{1/2}$(near equilibrium) might not be the ideal one.

In this case, model-dependent methods are e.g. using the before described Bell formalism (see equation 2.3)). Here the folding/unfolding rate at zero force $k^0_{f/u}$ and the corresponding distance to the folded and unfolded transition state $\Delta x_{f/u}$ can be obtained from a linear regression in a semi-log plot of the (un)folding rate log $(k_{f/u}(F))$ vs force. However this model neglects the fact that the applied force also effects the position of the transition state and therefore the rate extrapolated to zero force could tend to be unreliable, when the rate/force dependence becomes non-exponential [157].

For that reason many effort has been done during the last decade to improve the models for reconstructing the energy landscapes of a protein from single molecule force spectroscopy measurements [155].

So far the estimated distance to the unfolded transition state Δx_u for proteins ranges from a few angstroms (e.g. ubiquitin [116], SH3 [145] or Csp [128]) to several nanometers (e.g. GCN4 [142] or calmodulin [158]). The physical interpretation of this distance Δx_u resulted in the question if a thermodynamically stable protein could have stable state, which is tens of nanometers away from its folded state? Fernandez et al described by using theoretical simulations that the observed hopping during low force constant force and constant velocity measurements is related to a hopping around an entropic barrier rather than a hopping around the true folding/unfolding barrier of the protein [159, 160]. I.e. the observed hopping occurs rather between a total extended state and a collapsed state than between the folded (native) and the unfolded state of the protein. As a consequence any determination of the free energy ΔG from single molecule force spectroscopy measurements near equilibrium would not be relevant to bulk equilibrium measurements. However an alternative interpretation of this observation was proposed by Dudko et al [161], wherein the measured force dependent kinetics at low forces can indeed still probe the true zero-force barrier. Here the usage of a 2D energy landscapes taking a second reaction coordinate Q (fraction of native contacts) additional to the end-to-end distance x, reveals that depending on the chosen reaction coordinate x, the corresponding result can end into the interpretation from Fernandez et al [159, 160].

2.5 Steered Molecular Dynamics Simulation

Steered molecular dynamics (SMD) simulation describes the conduction of an all-atom molecular dynamic simulation, but with the application of an external mechanical force on the protein structure [162]. This can be achieved through the addition of a mechanical force parameter to the conventional force fields used. Depending on the value of applied

mechanical force the protein unfolds in a reasonable amount of calculating time. SMD simulation can simulate both constant velocity and constant force AFM experiments [162] and are therefore an ideal tool of choice as they can reveal useful insights, which can help to describe experimental data.

Many positive examples exist, where SMD simulation was used to interpret SMFS experiments [163]. For example with the help of SMD simulations, the observed hump in the experimental data from pulling on the I27 domain from titin with a constant velocity could be resolved. Therein so-called mechanical clamp could be identified in the I27 domain. Furthermore when mutating the corresponding residues the observed intermediate was vanished [134]. This is an excellent example where computational simulation and experimental data can both be used to solve and interpret the unfolding behavior of a protein under a mechanical force. Further examples of positive studies were done with ubiquitin [164] and cohesin I [123].

One of the more recent advances in SMD made it possible to measure the unfolding pattern of a polyprotein of six Ig domains from titin [165]. Observing the typical sawtooth pattern when the Ig unfold one after another. This simulation was done at a speed of 0.25 m/s to be able to observe unfolding in a reasonable amount of calculation time. Hence SMD simulation still need to be conducted with a velocity with a magnitude up to 10^6 times higher than velocities used in the single molecule force spectroscopy experiments. This results in a high discrepancy between simulated and experimental measured unfolding forces, e.g. in the case of Ig domains from titin up to 1000pN compared to 100-200pN of unfolding force. Therefore for a long time there was a gap of 2-3 magnitudes in unfolding force derived from SMD simulations and experiments.

From the simulation side, the new increased calculation capacity of conducting MD simulation as discussed in the previous chapter, should overcome the gap of experimental measured unfolding forces and simulated ones. However no recent conducted SMD simulation reaching the ms-s time range has been reported so far.

From the experimental side, Rico et al [166] used the advances from the high speed atomic force microscope from Ando [102], to construct a high speed force spectroscopic AFM, wherein he was able to apply high velocities of mm/s to the titin in order to close the discrepancy gap of simulations and experiments successfully.

Chapter 3

Proteins of study and research objectives

In the following the three proteins of investigation (BBL, Csp and gpW) will be discussed in detail, starting with its aminoacid sequence, topology and its role in the physiological environment. Furthermore a short review will be given of existing protein folding description and/or mechanical unfolding on each of them. However before, a short introduction to the used titin I27 domain and the ubiquitin protein will be given, as they were used throughout the thesis acting as a molecular fingerprint in order to interpret the experimental data. Afterwards the research objectives will be summarized.

3.1 The titin I27 domain and the ubiquitin protein

SMFS experiments using the AFM started quite from the beginning with the investigation of titin I27 and ubiquitin, as they are naturally occurring polyproteins and reveal a very stable and repeatable unfolding pattern. The 3D structure of both proteins is shown in figure 3.1.

The titin protein, a giant elastic protein ($\approx$ 40.000 aa) in the sarcomers of cardiac and skeletal muscles is by far the best known molecule by AFM force spectroscopy [87]. Around 90% of titin is made of globular repeating domains. The first AFM study was performed on these individual globular titin domains (e.g Immunoglobulin Ig) [105]. Afterwards different parts of titin were investigated, therein especially the I27 domain of titin was studied into detail [134, 131, 167]. This domain was chosen as its structure [168] and stability [169] was well known. The I27 domain consist of 89 aa and of a typical beta sandwich formed by two four-stranded sheets (see also figure 3.1 A). Later the idea of inserting the protein

of study between repeating I27 domains revealed an important advance in identifying the unfolding pattern of the sandwiched single protein [120, 128, 170]. AFM constant velocity measurements on the titin I27 domain at 400-600nm/s revealed an unfolding force of (204 ± 26)pN and a value of ΔL value of (28.4 ± 0.3)nm [167]. Force clamp measurements revealed a unfolding step size for I27 of around (24.5 ± 0.5)nm [150]. A recent study showed the refolding of the slow folding I27 domain ($k_f^0 \approx (4.9 \pm 0.6)s^{-1}$) when subjected to a small force below 5pN for several minutes using magnetic tweezers [171].

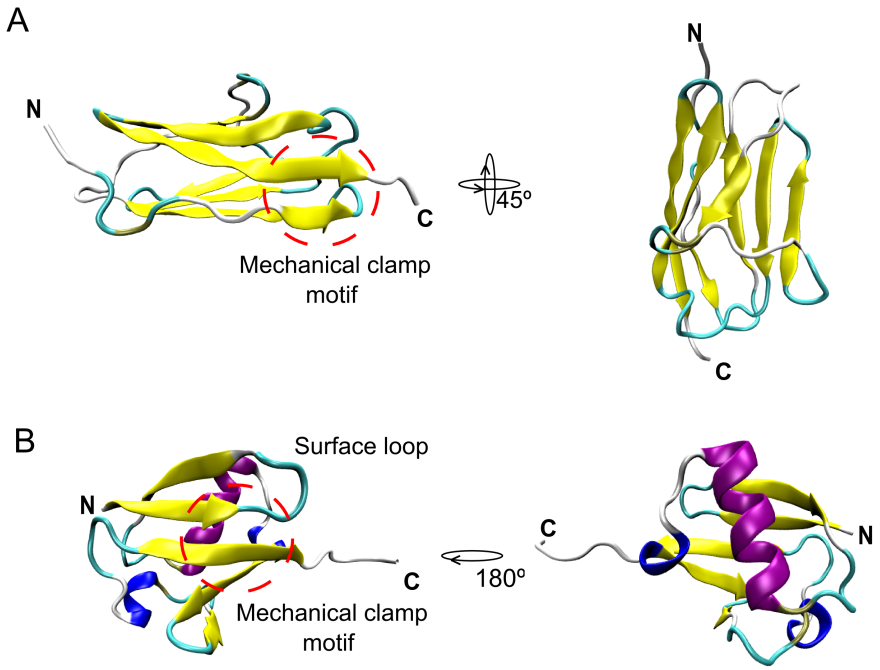

Figure 3.1: (A) The 3D structure of the I27 domain is illustrated revealing its eight β strand forming a barrel like structure. The mechanical clamp motif consisting of the two parallel β-strands is shown. (B) The 3D structure of ubiquitin depicts its five β strands and one α helix. The mechanical clamp motiv is illustrated and the surface loop, which was used to insert the protein of study is indicated [174].

Ubiquitin is a naturally occurring polyprotein of nine identical repeats and consists of 76 amino acids forming a characteristic α-β fold. Its folding and unfolding behavior has been studied in detail using chemical denaturants [172]. Ubiquitin is involved in protein degradation and other signaling pathways [173]. The mechanical behavior of ubiquitin is well reported in SMFS studies using the AFM in force clamp and constant velocity mode [116, 150, 164]. Furthermore it has been used to design the so-called pFS (plasmid for force spectroscopy) vector for performing constant velocity AFM measurements [174], which was used also in this work and will be described in more detail in the following Materials and Methods section. The basic idea of this vector is to include the protein of study into a surface loop of one ubiquitin repeat (called the ubiquitin carrier). This surface

loop is also indicated in figure 3.1 (B). In this study the unfolding force of ubiquitin at a constant velocity of 400nm/s was estimated to be around (189 ± 37)pN and the typical ΔL value to be around (23.4 ± 0.5)nm.

3.2 The BBL domain

BBL is the periphal subunit binding domain (PSBD) of the 2-oxogluterate dehydrogenase (2-OGDH) multienzyme complex of Escherichia coli, which structure has been resolved using NMR spectroscopy [175]. This multienzyme complex 2-OGDH consist of 3 subunits termed E1 (2-oxogluterate decarboxylase), E2 (dihydrolipoamide succinyltransferase) and E3 (dihydrolipoamide dehydrogenase) and is part of the primary energy-producing pathways of glycolysis and tricarboxylic acid circle [176]. Here the PSBD (BBL in the case of Escherichis coli) is reported to be involved in regulating and coordinating the interplay of channeling substrates between the core E2 and its subunits E1 and E3. The PSBS is a integral part of the E2 core, which bind to the pherical E1 and E3 subunits. The BBL structure used throughout this work consists of 50 residues, which is the same sequence used for the recently conducted FRET experiments [84]. Its topology consists of two short parallel alpha helices connected by a helix like turn (α2), an unstructured loop and a hydrophobic core, see figure 3.2.

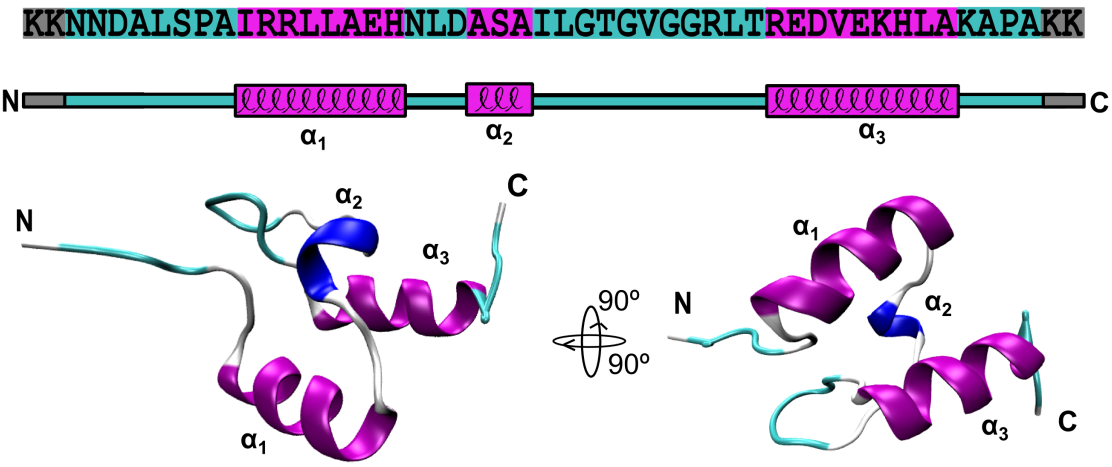

Figure 3.2: The used sequence of BBL, its topology and 3D structure are shown. The pdb file is 2WXC, but with an added tail PAKK at the C, a changed tail at the N terminus (KK→GSQ) and a changed aminoacid H18→W18

The PSBD (BBL) is postulated to be an important part inside the swinging arm mechanism as it binds to the E1 and E3 subunits [177]. As it has been mentioned before, the

BBL domain has been identified as a global one-state downhill folder using a variety of spectroscopic (both equilibrium and kinetic) probes and calorimetry [76, 77] and by high resolution equilibrium experiments using NMR [83] folding in microseconds via a marginal energy barrier. Using FRET measurements the downhill folding behaviour of BBL could be recently observed by the movement of the single peak in the FRET histograms during chemical denaturation [84]. Another experiment revealed the effect of salt on the BBL, inducing complete refolding of acid-denatured BBL into the native structure and therefore modulating the folding cooperativity through electrostatic screening [178]. Furthermore the sensing capacity of BBL to the pH (pH range 7-3) and ionic strength was shown to be useable as an application for engineering ultrafast biosensors [179]. However the pH dependence (pH range 6-12) of BBL stability, folding kinetics and mechanisms are quite puzzling and were recently resolved in detail [180]. Therein the observation of protonated and unprotonated species of one of its residues of BBL is explained to be the reason for observing two inter-converting species of BBL during chemical denaturation between a pH range of 7-8. This is thought to be the main cause of observing different (un)folding behavior of BBL (downhill vs 2-state) using absorbance, fluorescent and NMR spectroscopy [181] and single molecule FRET within different research groups [84, 182].
In sum, BBL is described as a one-state downhill fast folding protein with kinetics in the μs range ($\approx$20μs at room temperature) and having a truly negligible (non existent) energetic barrier (<1RT). BBL unfolds gradually by populating a single conformational ensemble, which amount of native structure is proportional to the level of denaturing stress.
This together with its physiological role inside the above mentioned swinging arm mechanism inside the 2-OGDH multienzyme complex [177] let to description of BBL to be an example of a molecular rheostat or a nanospring [76, 179].
Thus single molecule force spectroscopic measurements on BBL is suggested to reveal new insights into the behavior of a molecular rheostat under a mechanical force

3.3 The cold shock protein B (Csp)

Cold shock proteins (Csps) are a subgroup of the cold-induced proteins expressed preferentially in bacteria and other organisms when the growth temperature drops to regulate the adaptation to cold stress. However, they are also present under physiological conditions to regulate other biological functions. All Csps bind single-stranded RNA and DNA through two strutural well conserved motifs (called RNP1 and RNP2). Nine members of the Csp family have been identified so far, named in alphabetical order CspA to CspI. In

the case of a temperature drop, Csps are thought to bind RNA or DNA to avoid the formation of secondary structures, which hinders transcriptional (and translational) processes [183]. This function of the Csps is then often called RNA-chaperone function. The CspB from the hyperthermophilic bacterium Thermotoga Maritima (from hereon just called Csp throughout this work) the protein of study consists of 66 residues. Its topology consists of six antiparallel β-strands forming a β-barrel (see figure 3.3), which has been resolved using NMR spectroscopy [184].

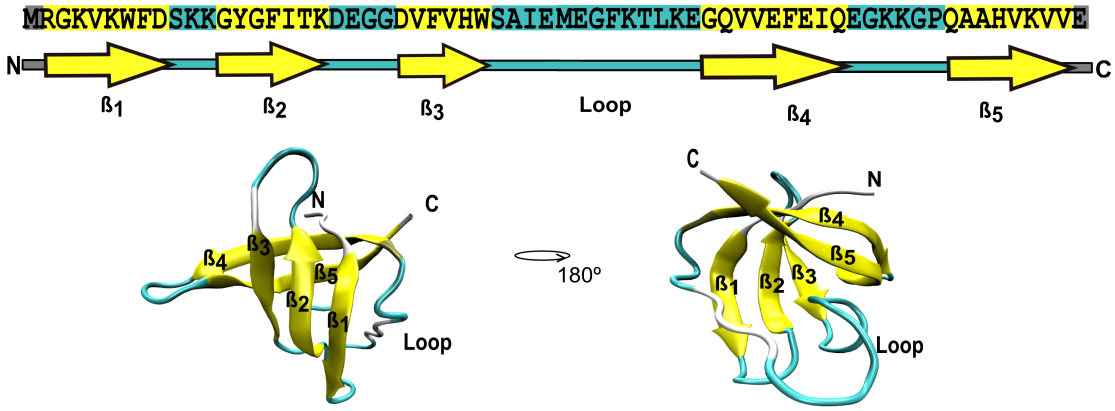

Figure 3.3: The used Csp sequence, topology and 3Dim structure are shown. The pdb file of Csp is 1G6P

Equilibrium measurements using calorimetry [185] describes the thermal unfolding of the TmCsp as a clear 2-state process and it reaches its maximum thermal stability (ΔG = 6.5kcal/mol, ~11RT) at 30°C at a pH 7 with a corresponding melting temperature of 82°C. The kinetics were measured using a stop-flow spetrometer [186] and revealed a fast refolding rate of around 550 s^{-1} and a slow unfolding rate of 0.02 s^{-1} at pH 7 and 25°C. At a single molecule level the TmCSP is the most extensively studied protein using FRET fluorescence [73, 67] and recently also by AFM force spectroscopy [128, 187], both revealing a clear two-state folding behavior. Constant velocity AFM measurements revealed an unfolding force of the TmCSP to be around 78pN at a velocity of 400nm/s and a difference in contour length ΔL of 23.5nm, therefore indicating a high mechanical stability.

In sum, the TmCSP is well described two-state protein especially also within single molecule measurements. Additionally it is a fast folding protein (ms range) and shows sufficient mechanical resistance. That made it a protein of choice during this work to have a comparison to the mechanical force measurements done with the BBL protein.

3.4 The gpW protein

The gpW protein plays a crucial role in the head-to-tail assembly of the *Eschericha coli* bacteriophage λ. During head morphogenesis of bacteriophage λ syntesis gpW is suggested to polymerize into a ring-like structure and its exact position has been recently imaged using electron microscopy inside the head-to-tail connector [188]. It is further described to interact to two other proteins forming ring-like structures inside the head-to-tail connector, called the portal protein gpB and the head completion protein gpFII. Additionally gpW has been reported to stabilize DNA packaging inside the head of the bacteriophage λ [189]. The gpW protein used throughout this work contains 65 residues and recent NMR structural studies have reviewed that gpW consists of two antiparallel α-helices and a single antiparallel two-stranded β-sheet (see figure 3.4) [190].

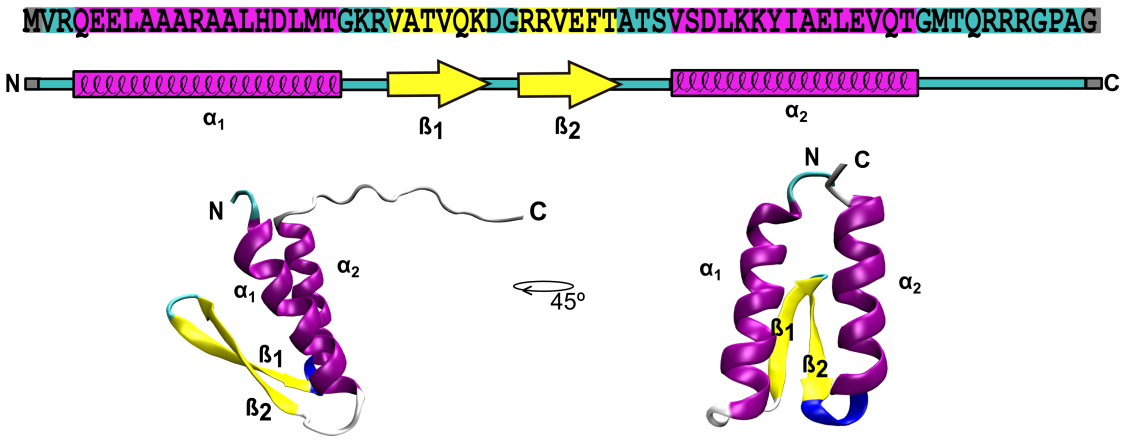

Figure 3.4: The used gpW sequence, topology and 3Dim structure are shown. Compared to the pdb file 2L6Q, 3 more resedues were added in the unstructured C-terminus

Conducted thermodynamic measurements revealed at a first glance a two state folding mechanism of gpW over a free energy barrier ΔG of 5RT [191]. However more detailed multiprobe thermodynamic, calorimetric and kinetic studies of gpW showed the typical behavior of a downhill folding protein [78]. Thereafter gpW folds in the microsecond range over a marginal free energy barrier (about 1 RT), being out of the global downhill folding regime but showing a large complexity during unfolding. This complex unfolding behavior was visualized with an atom-by-atom analysis with of gpW using NMR spectroscopy [35]. As mentioned before the description of a complex free energy landscape of gpW was further strengthened through a NMR relaxation experiment, revealing a complex free energy landscape [79]

Thus the gpW protein was chosen to study with single molecule force spectroscopy as it

folds/unfolds in the microsecond range in the downhill regime close to the global downhill limit, but it lacks of a complicated dependence of pH value compared to the BBL protein.

3.5 Research objectives

The research objectives consist of investigating the three before in detail discussed fast folding proteins BBL, Csp and gpW under a mechanical force both experimentally with SMFS using the AFM and computationally using SMD simulations. The three small proteins were chosen as they cover a wide range in folding time in the fast folding regime from ms to μs range, but also show different folding behaviors from two-state over nearly downhill to global downhill (one-state). Furthermore all three proteins show a different topology from all β to $\beta + \alpha$ and all α.
The following table 3.1 gives an overview of the mentioned properties for the three investigated protein samples BBL, Csp and gpW.

	pdb file	# residues	topology	folding behavior	folding time	energy barrier
BBL	2WXC	50	α	global downhill	μs	<1RT
Csp	1G6P	66	β	two-state	ms	∼11RT
gpW	2L6Q	65	α+β	nearly downhill	μs	∼1RT

Table 3.1: Overview of structural properties and folding behavior of the three proteins used BBL, Csp and gpW.

Thus the mechanical description of the three fast folding proteins of study following different types of folding mechanisms is thought to give further insights into the field of protein folding.
The thesis is now organized into the following chapters 4-7. The materials and methods used to build and purify the protein sample, to perform SMFS experiments using the AFM, to conduct the data analysis and finally to execute stochastic kinetic and SMD simulations will be in detail presented in chapter 4. Within chapter 5 to 7 the results for each protein of study will be presented in chronological order from the starting of the project to its finalization.

Part II

Materials and Methods

Chapter 4

Materials and Methods

This chapter explains the whole development of the project, from the initial idea to the final analysis of the experimental data and is divided into four main sections. The first section describes the molecular biology methods used to construct and obtain the protein sample. Then the second section explains the experimental setup of the two used atomic force microscopes. The third section will give the details about the methods and software used to analyze the experimental data and finally the forth section will describe the performed Steered Molecular Dynamics (SMD) simulations.

4.1 Molecular biology

For conducting single molecule force spectroscopy (SMFS) measurements on the Csp, gpW and BBL protein containing molecular constructs, a high importance lies in the DNA manipulating, cell expression and protein purification methods to produce the final polyprotein sample. In general, the better the purity and quality of the protein sample the better the final AFM traces. Throughout this work many different approaches and methods have been tried until the ideal one was found and finally used for all the polyprotein constructs.

4.1.1 Polyprotein synthesis

In order to obtain a DNA plasmid containing the polyprotein (gene) of interest, a strategy was used, which is based on a oriented and stepwise insertion of DNA fragments into a suitable expression vector.

The basic idea of this approach consists of using cutting enzymes with compatible cohesive ends and which ligation product generates a new restriction site that cannot be cut by neither one of them. Therefore, this strategy allows a consecutive addition of genes (DNA

fragments) into the plasmid vector and results in the designed polyprotein construct. The cutting enzymes *Bam*H I and *Bgl* II have exactly this property and were used throughout this work. Here, the new restriction site generated after the ligation of *Bam*H I and *Bgl* II cannot be cut by neither *Bam*H I or *Bgl* II digestion. Before using this strategy each new protein sequence i.e the Csp, BBL and gpW encoding sequence needed to be revised for the restriction sides of *Bam*H I and *Bgl* II, in order to confirm that both cutting enzymes could be used. The corresponding maps and sequences of the used plasmids vectors (pRSETa and pBAT4) can be reviewed at https://www.embl.de.

The discussed insertion strategy of a gene (DNA fragment) into the plasmid vector basically consist of four steps, which are illustrated in figure 4.1.

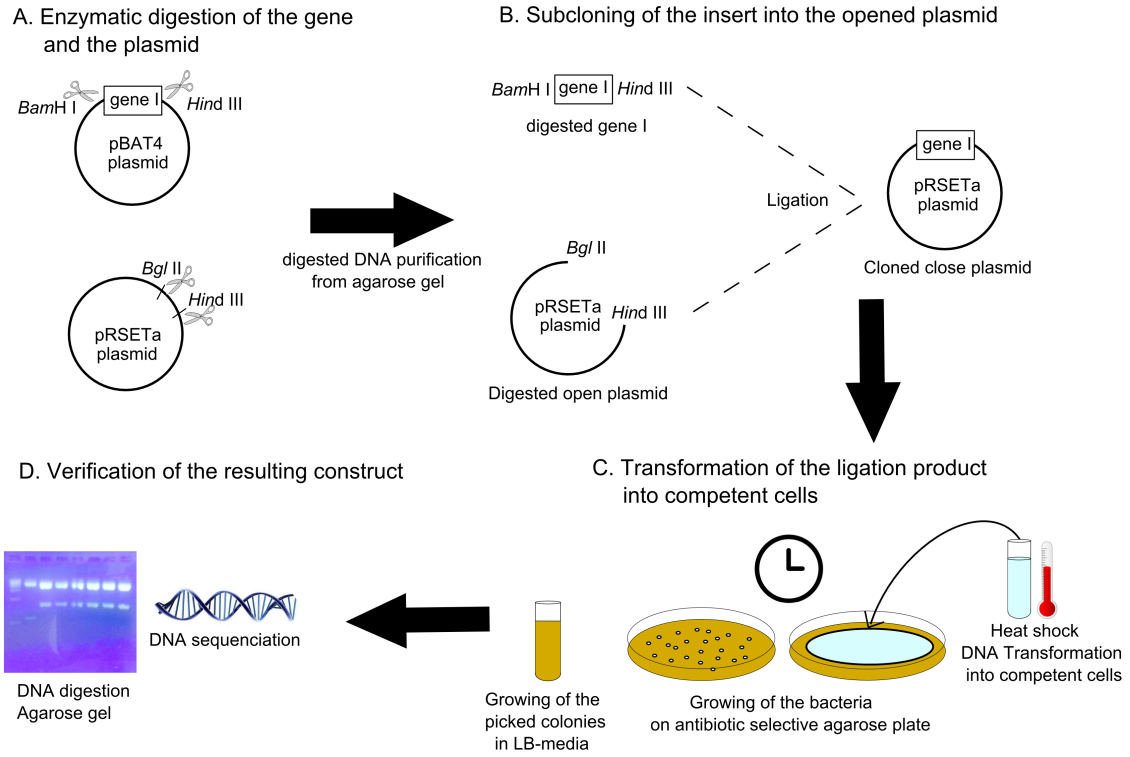

Figure 4.1: The cloning strategy used for the polyprotein synthesis. A) Cutting of the gene and the plasmid vector using specific cutting enzymes. B) Ligation of the gene and opened plasmid vector. C) Transformation of the ligation product into competent cells D) Verification of the resulting polyprotein construct

A) Enzymatic digestion of the gene and the plasmid:

Here, the gene I containing plasmid (pBAT4) and the plamid used for the polyprotein construction (pRSETa) were extracted from the grown cells (LB-media (+0.1% 100mg/ml Ampicilin); 37°C overnight) using a DNA-plasmid extraction kit following the manufacturer protocol. Corresponding kits from Qiagen (http://www.qiagen.com) or Machery-

Nagel (http://www.mn-net.com) have been used throughout this work. Purified plasmids were eluted in sterile water and stored at 4°C in the fridge.

After purification, the concentration of the plasmids was estimated by absorbance by either the nanodrop system (http://www.nanodrop.com/) or a UV-spectrometer (Cary Eclipse Agilent, http://www.agilent.com). If not otherwise mentioned all digestion enzymes used in this work were from Invitrogen (http://www.lifetechnologies.com). In general, the final digestion reaction solution was adjusted to a final volume of 25μl with sterile milipoore water and kept for 1 hour in the oven at 37°C. The amount of each reactant used during the digestion was adopted from the included manufacturer protocols, wherein one unit of enzyme is usually required to digest 1μg of DNA in 50μl of the reaction mixture in one hour at the given reaction temperature. Each reaction contained around 4μg of DNA plasmid.

Here both plasmids (one containing the gene insert and one containing the polyprotein vector) usually were double digested at one time. Afterwards the final reaction solution was treated with 0.01 units of Alkaline Phosphatase (CIAP from Promega http://www.promega.com) to prevent self ligation of the plasmid vector. The CIAP reaction solution was kept for another 30 minutes at 37°C in the oven.

After digestion, the gene I insert and the opened plasmid vector were screened and purified from DNA Agarose gels (Low EEO, Conda, http://www.condalab.com) in TBE buffer. For the opened plasmid vector (>1000bp) a 1% Agarose gel was produced whereas for the DNA gene (>100bp) a 2% Agarose gel was needed. For DNA staining SYBR® Safe DNA Gel Stain (Thermo Scientific, http://www.lifetechnologies.com) was added. The DNA Agarose gel was run using a BioRad equipment http://www.bio-rad.com.

Afterwards, the corresponding bands were cut out of the DNA gel and the DNA was extracted from the agarose gel, with extraction kits again from either Quiagen or Machery-Nagel. The final solution was typically diluted into 40μl sterile milipoore water and stored at 4°C until further use.

B) Subcloning of the insert into the opened plasmid:

For the ligation T4-Ligase (Thermo Scientific) and its corresponding manufacturer protocol was used. Briefly, a 1:10 Mol ratio was adjusted between the amount of the plasmid vector and the DNA gene insert. Therein, 50ng of plasmid vector were used in every reaction and the corresponding amount of gene was then calculated and added according to the following formula which puts the number of basepairs (bp) and the molar mass g/mol into

relation:

$$\text{insert (ng)} = 10 \times \frac{\text{plamid vector (ng)} \times \text{ insert (bp)}}{\text{plamid vector (bp)}}$$

The ligation was carried out at 16°C overnight.

C) Transformation of the ligation product into competent cells:

Typically, DH5α heat shock competent cells (e.coli; Invitrogen) were used to facilitate DNA transformation and DNA cell conservation at -80°C. The protocol for the DNA plasmid transformation by a heat shock was as follows:
Before the heat pulse the competent cells were unfrozen but kept in ice. Then 50μl of competent cells were transferred into sterile and prechilled 1.5ml eppendorfs. Here typically, 1μl of the ligation solution was added. A positive (uncut DNA plasmid) and a negative (no plasmid) control was also made in parallel. All the samples were then kept on ice for around 30 minutes for the DNA to be adsorbed on the cell membrane. The so-prepared samples were hold exactly 40 seconds in the water bath or heating block at 42°C. After the heat pulse the samples were put back on ice and chilled down for around 2 minutes. Thereafter, 450μl of preheated (42°C) and sterile LB was added to the cells and put in the incubator at 37°C for one hour.
Then the samples were centrifuged for 5 minutes at around 13000 rpm in a microcentrifuge. Afterwards the cell containing pellet was redissolved in 100μl of sterile LB. Typically, 1/10 culture dilutions were then spread onto LB-agar+Ampicilin plates. The so prepared plates were incubated around 15 hours at 37°C. in the incubation oven at 37°C overnight for growing which means for around 15 hours. In the positive event of grown colonies the plates were stored under sterile conditions at 4°C until use.

D) Verification of the resulting construct:

Single colonies were the isolated and grown in LB media (+0.1% 100mg/ml Ampicilin) at 37°C overnight. Before the DNA plasmid was extracted from the cells 1ml volume of the cells were mixed with sterile glycerol stock (50% glycerol/50% LB), snap frozen in liquid nitrogen and stored in the -80°C fridge.
Then the extracted plasmid was sequenced (Secugen or Parque Scientifico, Madrid, Spain) using the T4 polymerase. Additional oligonucleotides used for sequencing are will be listed in the corresponding section (e.g see table 4.7)
Here additionally, the correct length of the polyprotein construct was always confirmed by enzymatic digestion screened with a DNA agarose gel. In the case of the confirmation

of successful cloning, the four steps A-D were repeated consecutively until the desired full polyprotein was constructed.

4.1.2 Cell lysis, protein expression and purification

In general, the protocols for cell lysis, protein expression and purification were adopted from the ones used in the Marriano-Carrion lab (see [174]). Especially the cell lysis protocol has been found to produce the polyprotein samples best. Its process is less harsh to the protein construct compared to other protocols using e.g. sonication and the french press, which had been also tried but discarded finally.

Protein expression

For protein expression the for the desired polyprotein construct encoding DNA plasmids (pRSETa) were transformed into C41 competent cells (e.coli, Novagen, http:www.novagen.com) using the previous explained transformation protocol (see section 4.1.1). C41 cells are especially designed for the expression of toxic proteins. As the polyprotein structure has a toxic pattern for regular expression cells the usage of the C41 turned out be beneficial for the expression of the used polyproteins.
The following conditions for protein expression were used for every built polyprotein construct, which have been shown to work successful for a lot of polyprotein constructs (see e.g. reference [174]).
The C41 cells carrying the polyprotein construct were usually grown from a low volume (≈2ml) and scaled up to a final volume of 1l of LB containing the adequate concentration of ampicilin (+0.1% 100mg/ml). This small volume (nucleus) was grown for around 6 hours at 37°C and 250rpmm agitation in the incubator. Afterwards the 2ml volume was put into 50ml of sterile LB ampiciline inside a 500ml bottle and left for growing at 250rpm at 37°C.
Then the overnight culture was put in a final volume of 1l of LB ampiciline inside a 5l flask in order to have sufficient amount of oxygen to improve the growing of the cells. The cultures were then incubated at 37°C at 180rpm typically for 3-4 hours until an optical density (OD 600nm) value of 0.6-0.8 was reached, which was measured in time intervals during the growing process using a UV-spectrometer. Once the OD was reached, a volume of 0.1 % IPTG (Isopropyl-β-D-thiogalactopyranosid) at a concentration of 100mg/ml was added to the grown cells in order induce the protein expression by T7 promoter activation. Therefore the cells were incubated further for another 4 hours at 37°C.

Afterwards the cell culture was transferred into 500ml centrifuge tubes for the F-10 rotor and centrifuged at 8000rpm for 30 minutes at 4°C. Then the supernatant was discarded and the pellet (typically 2-3g of weight) was transferred into 50ml falcons and stored overnight in the -80°C freezer to start the cell breakage (lysis) and to avoid the cell protease to digest the expressed polyprotein construct.

Cell lysis

The cell lysis process starts from the frozen cell pellet after its resuspension by vortexing in 40ml of chilled 1x Binding buffer used for the HisTag affinity column purification (Sodium phosphate pH7 50mM, NaCl 500mM, Imidazole 50mM and TCEP 1mM; see also table 4.2). Here one Protease Inhibitor Cocktail tablet (Roche http://www.roche.com) was added and vortexed further. During the resuspension the sample was kept on ice in a 50ml falcon until pellet was well dissolved and the solution became homogeneous milky. Additionally. at this point 50mM of Imidazole was added here in order to equilibriate the sample before the injection into the HisTag column. The dissolved pellet was then once snap frozen in liquid nitrogen and once snap thawed at 42°C using a water bath until a small ice-ball remained in the falcon.
Then the lysis was continued by adding 40mg lysozyme, which was previously dissolved in the prepared lysozyme buffer (see table 4.1) and stored on ice for about 30min. Then the samples were further put in the cold room at 4°C and left on a gently shaker for another 30min. Here the solution typically became more inhomogeneous and dense.
Afterwards 4ml of Triton X-100 solution (10% w/v), 0.2mg of DNASE I (dissolved in the corresponding DNASE I buffer) and 0.2mg of RNase A (dissolved in the corresponding RNASE A buffer) were added to the solution and left for further incubation on the shaker in the cold room for another 60-90 minutes until the lysate was completed.
The lysate was then centrifuged using the SS-34 rotor at 18.100 rpm for 30 minutes. After the centrifugation the supernantant was filtered through 0.45μm filter (Milipoore, http://www.merckmillipore.com) before its injection into the HisTag affinity column. The clarified lysate could now be directly injected for the HISTAG affinity column to start the purification process.

Protein purification

High performance liquid chromatography (HPLC) (Agilent, http://www.agilent.com) was used for the polyprotein purification in two steps. The first purification was done with

buffer	content	amount
Lysozyme	Tris-HCl pH 8.4	10mM
DNASE I	Tris-HCl pH 7.5	20mM
	Glycerol	50% w/v
	$MgCl_2$	1mM
RNASE A	Tris-HCl pH 7.5	10mM
	NaCl	15mM

Table 4.1: Overview and content of the used buffers during the cell lysis process

a nickel-affinity HisTrap column (Ge Healthcare, http://www.gehealthcare.com). This type of column can be used as each polyprotein construct contains a HisTag made of 6 histadine at the N terminus of the protein which specifically binds to the Ni loaded nickel-affinity column (agarose beads). In changing the concentration of imidazole of the elution buffer the polyprotein sample is than washed out from column. The second purification is performed with a size exclusion Superdex 200 column (Ge Healthcare). Here the proteins are separated by its size and larger proteins will run through the column faster than smaller as they will be to diffuse into the pores of the column matrix.
For the first purification using the HPLC HisTrap column the following binding buffer and elution/washing buffer were prepared (see table 4.2).

Buffer	Binding	Elution
Sodium Phosphate pH 7.0	50mM	50mM
NaCl	500mM	500mM
Imidazole	50mM	500mM
TCEP	1mM	1mM

Table 4.2: Content of the used buffers during the HisTrap column purification

After filtering the produced buffers the HPLC was run with the following program at a rate of 2ml/min (see table 4.3). 10ml of the sample (lysate) was injected at once into the column, which was repeated 4 times before the HPLC switched from the binding buffer to the elution buffer in order to wash out the bounded polyprotein from the column.
After the first purification normally a SDS protein gel was run in order to confirm the eluted protein amount, detected by the 280nm UV signal of the HPLC. Then, typically ultrafiltration Amicon 3k filters (Milipoore) were used to dilute and exchange the elution buffer with the washing buffer used for the superdex size exclusion column (see table 4.4).

Min	Binding Buffer %	Elution Buffer %
0 - 80	100	0
80 - 86 (gradient)	0	100
86 - 126	0	100
126 -136	100	0

Table 4.3: Typical program used for the HisTrap column purification

Buffer	Superdex elution
Buffer Phosphate pH 7.0 NaP	50mM
NaCl	300mM
TCEP	2mM

Table 4.4: Washing buffer content used for the superdex size exclusion column

Afterwards, the samples were concentrated usually to a final volume of 2-3ml, depending on the amount of the expressed protein from the HisTrap purification.
During the second HPLC step with the superdex 200 size exclusion column 0.5ml could be injected into the column for each run. The protein was then eluted with the superdex washing buffer for 60min at a speed of 0.5 ml/min. Afterwards again a SDS protein gel was used to identify the protein in the corresponding fraction.
Then again ultrafiltration Amicon 3k filters were used to dilute and exchange the superdex buffer with the final measuring buffer which was normally 1x PBS at pH 7.4. The final concentration of each sample was roughly estimated using the nanodrop 280nm absorbance function with an extinction coefficient and a molecular weight calculated for each of the polyprotein constructs (see http://www.biomol.net/en/tools/proteinextinction.htm). The corresponding concentration was around 1-3mg/ml for each polyprotein sample. The samples were then distributed equivalently into 200μl fraction tubes, which were than snap frozen with nitrogen and put in the -80^{o}C freezer for further storage.

4.1.3 Production of the titin I27 domain sample

The pRSETa plasmid vector (see [193]) was used for all polyprotein constructs containing the I27 titin domain (for sequence information see [167]) and was a kindly gift from the Dr. Mariano Carrion lab (http://carrionvazquez-lab.org). The I27 domain was needed as it serves as a molecular fingerprint in the AFM measurements.
The received pRSETa vector already consisted of eight titin I27 domains. The tactic now

was to use the polymerase chain reaction (PCR) in order to copy and amplify a single titin I27 domain from the received sample. The amplified DNA gene encoding for the titin I27 domain was flanked by 5' BamHI and 3' HindIII-(-C-C-)-BglII reaction sides. The two cysteines (-C-C-) were put at the end of the I27 titin gene and serve as a linker to the gold surface. In order to conduct the PCR (BioRad) the following oligonucleotides (Sigma Aldrich, http://www.sigmaaldrich.com) were ordered (see table 4.7). Each solution of the oligonucleotides was adjusted to a final concentration of 125ng/μl and confirmed using the UV-spectrometer. The PCR solution reaction was prepared containing 50ng of the DNA template (I27 domain containing pRSETa plasmid), 125ng of each oligonucleotide, 10mM dNTP mix and 1 unit of Pfu DNA polyperase (Promega, http://www.promega.com) and filled up to 50μl using sterile milipoore water. During the preparation the PCR solution was kept on ice. Then the solution was spun down and the Pfu polymerase (Promega, http://www.promega.com) was added in the end. For the PCR program the recommended conditions from the Pfu polymerase manufacturer manual were followed. The table 4.5 shows the PCR program used for DNA replication.

Step	Block	Cycles	Temperature	Time
1	Initial Denaturation	1	95°C	30 seconds
2	Denaturation	30	95°C	30 seconds
	Annealing		55°C	45 seconds
	Extension		72°C	2 minutes
3	Final Extension	1	72°C	10 minutes
4	Soak	∞	4°C	1 minute

Table 4.5: PCR program

After the run cycles, the PCR solution was screened in a 2% DNA agarose gel and resulted in a distribution of bands. The lowest band was cut (≈300bp) as it corresponded to the expected size of a single I27 domain (267bp (3×89 aa)).

A second PCR was then run in order to amplify the concentration of the single I27 domain. Therein the same PCR solution as in table was prepared, but changing the pRSETa $I27_8$ (50ng) vector to the produced single I27 titin domain (50ng) as a template. Then again a 2% DNA agarose was run in order to check and to clean the resulting titin I27 monomer containing DNA gene. This produced 5' BamHI-I27-BglII-C-C-HindIII 3' DNA gene was now inserted into the pRSETa vector with the protocol as explained in figure 4.1 in section 4.1.1. Therefore the $I27_8$ needed to be cut out of the original receive pRSETa plasmid.

In the same way the titin I27 dimer ($I27_2$) and titin I27 trimer ($I27_3$) were constructed consecutively in adding another I27 monomer to the titin I27 monomer pRSETa plasmid. Figure 4.2 sketches the methods and used cutting enzymes to obtain the three titin I27 samples inside the pRSETa plasmid.

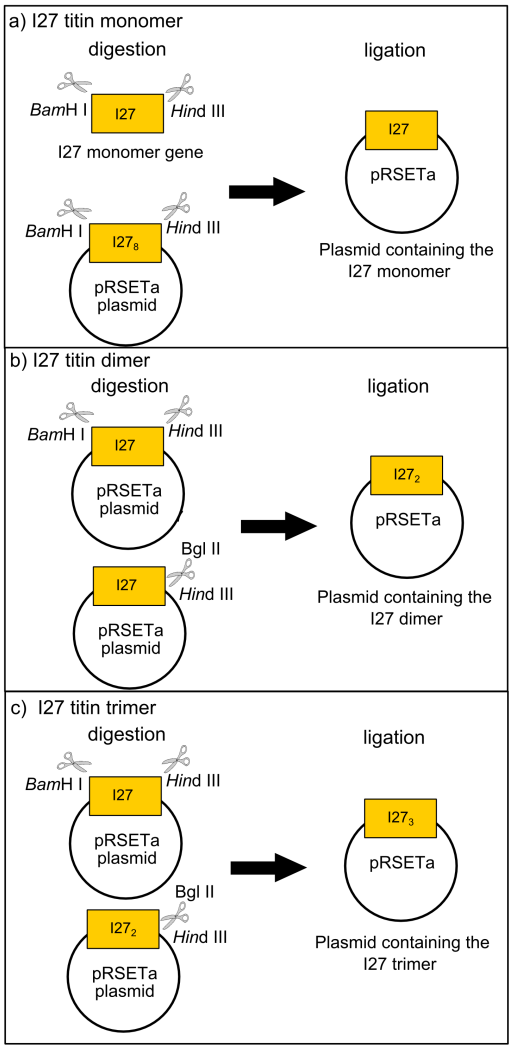

Figure 4.2: The cloning strategy used for the titin polyprotein synthesis. a) The insertion of the I27 monomer gene into the pRSETa plasmid vector b) Adding another I27 monomer to receive the I27 dimer inserted into the pRSETa vector c) Adding another I27 monomer to receive the I27 trimer inserted into the pRSETa vector

The correct sequence of each construct $I27_1$, $I27_2$ and $I27_3$ was confirmed by sequenciation using the T7 promoter primer. Within this protocol the inserted domains are always separated by the amino acids Arginine (R) and Serine (S), as they form the ligation product of the BamHI and BglII cutting sides. The produced single monomer, dimer and trimer of the titin I27 domain now allowed to start all the polyproteins designed in this work, which use the titin I27 domain as a molecular fingerprint.

4.1.4 Construction and purification of the single Csp, BBL and gpW domain containing polyprotein samples.

As each of the three polyprotein constructs was designed, built and purified under the same conditions, this section summarizes their construction together.

All three polyprotein samples were designed by flanking each side of the single Csp, BBL or gpW domain with three titin I27 domains in order to get a $I27_3$-Csp-$I27_3$, $I27_3$-gpW-$I27_3$ and a $I27_3$-BBL-$I27_3$ construct. The sequences of the Csp, gpW and BBL were all cloned inside a pBAT4 vector having a BamHI cutting side at the beginning and a BglII/HindIII cutting side in the end (each ordered from Clonetech, http://www.clontech.com/).

In starting from the produced $I27_3$ sample inside the pRSETa plasmid vector (see chapter 4.1.3; figure 4.2), the three constructs were built in parallel each of them consecutively in two steps, which is shown schematically in figure 4.3. During the first step the cloned protein sequences inside the pBAT4 vector and the $I27_3$ construct inside the pRSETa vector were digested with BamHI/HindIII and BglII/HindIII, respectively. Then the protein fragment (Csp, BBL or gpW) was ligated into the $I27_3$ vector to receive the $I27_3$-Csp, $I27_3$-gpW and $I27_3$-BBL construct. During the second step the $I27_3$ construct inside the pRSETa vector and the produced $I27_3$-Csp, $I27_3$-gpW and $I27_3$-BBL construct were cut with BamHI/HindIII and BglII/HindIII and ligated together in order to finally receive the three desired polyprotein constructs $I27_3$-Csp-$I27_3$, $I27_3$-gpW-$I27_3$ and $I27_3$-BBL-$I27_3$.

Each step was followed with a 2%-DNA agarose gel and sequenced to confirm the correct insertion of all three constructs. However, because the T7 promoter primer used for sequencing can just read until $\approx$ 800bp, specific primers were designed and ordered (see table 4.7). Here primer 1 Csp bound between the titin I27 domain and the beginning of the Csp domain and primer 2 Csp between the end of the Csp and the beginning of the titin I27 domain. Similar primers were designed and used for the BBL and gpW construct respectively (see table 4.7).

The final three constructs were now transformed into the C41 competent cells in order to start the protein expression. For the purification the same conditions as discussed in the general purification protocol (see chapter 4.1.2) were used for all the three samples. As the amount of purified protein was huge ($\approx$ 8mg) for all three constructs, the total amount was mostly divided into 4 parts and concentrated separately to a final volume of 3ml each.

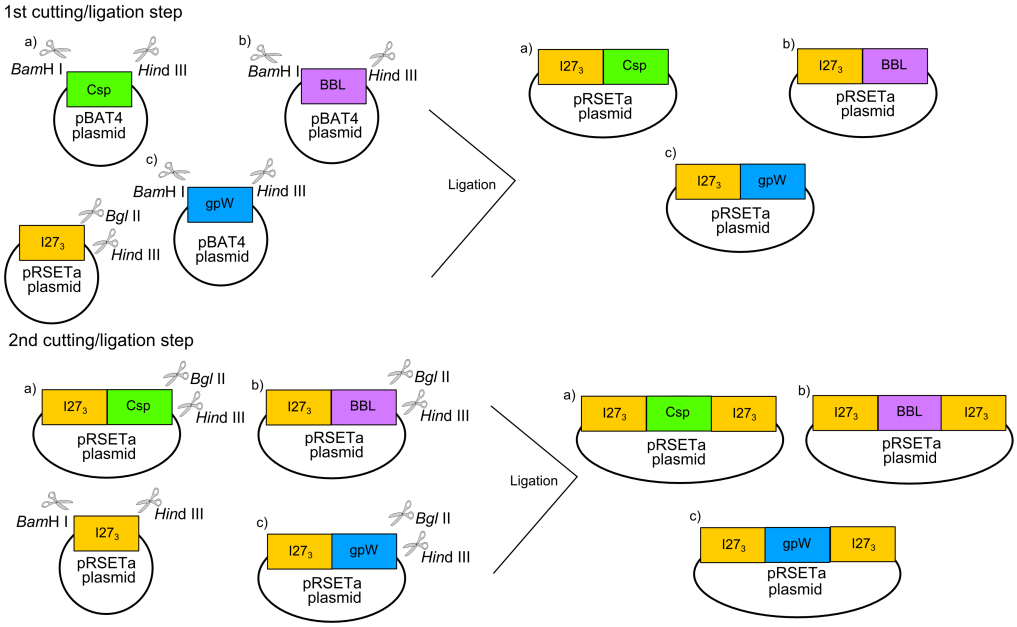

Figure 4.3: The schematic shows the approach used to build all the three samples containing a single domain of a) Csp, b) BBL and c) gpW respectively, flanked by three I27 domains on each side.

4.1.5 Construction and purification of the poly-BBL samples.

Two multiple BBL domain polyprotein constructs were designed. One was made of three BBL domains inside six titin I27 domains in alternating order $I27_2$-BBL-I27-BBL-I27-BBL-$I27_2$ an the other one was a pure poly-BBL construct, consisting of six BBL domains BBL_6.

The $I27_2$-BBL-I27-BBL-I27-BBL-$I27_2$ construct was build in four steps starting with the titin I27 monomer and the titin I27 dimer $I27_2$ inside the pRSETa vector (obtained in chapter 4.1.3).

During the first step the ordered BBL sequence inside the pBAT vector and the I27-monomer and $I27_2$-dimer construct inside the pRSETa vector were cut all in parallel with BamHI/HindIII and BglII/HindIII respectively. Then the BBL fragment was ligated into the I27-monomer and into $I27_2$-dimer vector to receive the I27-BBL and the $I27_2$-BBL construct. During the second and third step the I27-BBL gene was cut with BamHI/Hind III and two times repetitively ligated into the produced $I27_2$-BBL vector both produced in step one. Then in the forth and last step the resulting construct $I27_2$-BBL-I27-BBL-I27-BBL was then finalized by adding the titin I27 dimer $I27_2$. Figure 4.4 sums schematically the procedure used to build the $I27_2$-BBL-I27-BBL-I27-BBL-$I27_2$.

The second poly-BBL sample was built by adding one BBL to the another consecutively until the fully BBL_6 construct was finished, starting from putting one single BBL domain

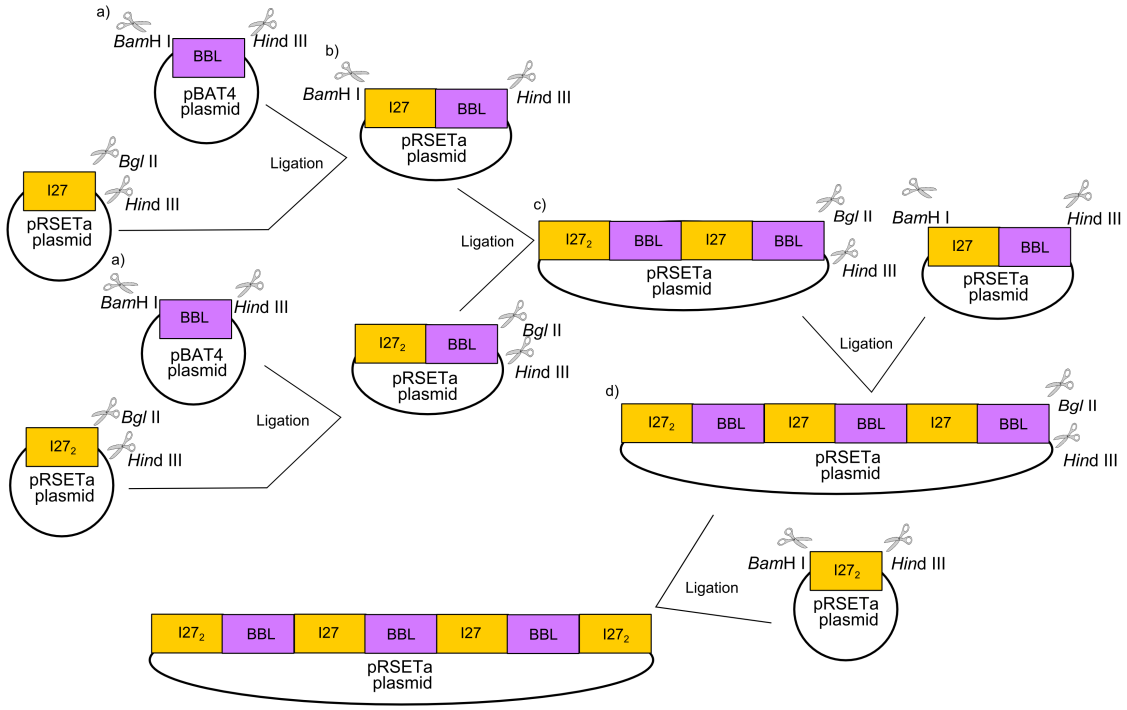

Figure 4.4: The schematic shows the approach used to build the $I27_2$-BBL-I27-BBL-I27-BBL-$I27_2$ in four steps, consecutively from a) to d).

in the pRSETa vector. Figure 4.5 shows the corresponding steps.

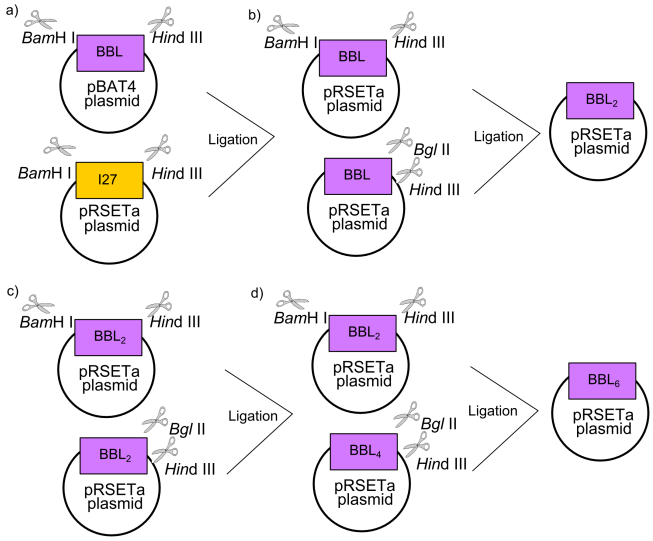

Figure 4.5: The sketch depicts the approach used to build the BBL_6 polyprotein construct from a) to d).

For both constructs each step was followed by a 2%-DNA agarose gel in order to confirm the correct insertion during the building process. Additionally each step was also sequenced. The final two constructs were then transformed into the C41 competent cells in order to start the protein expression. For the purification the same conditions as discussed in the general purification protocol (see chapter 4.1.2) were used for all the three samples. Also

here the amount of purified protein was in a similar range like before (≈ 6mg) for the two constructs.

4.1.6 Construction and purification of the BBL-monomer and dimer inside the pFS vector.

The pFS vector was a kindly gift from the Dr. Mariano Carrion lab (http://carrionvazquez-lab.org), where this vector was designed and built (see [174] for more information). This vector consists already of six ubiquitin domains. One of the six ubiquitin domain (the uniquitin carrier) carries an elongated surface loop, in which the protein of investigation can be inserted by using specific cutting enzymes. Here the ubiquitin domain like the titin I27 domain before acts as the molecular fingerprint.
In principle the pFS vector enables the unambiguous mechanical characterization of proteins especially with low mechanical stability or/and complex mechanical unfolding behavior. Here it was decided to probe this vector with a single BBL domain and additionally also with a BBL dimer (BBL_2) inside the surface loop of the ubiquitin carrier. Figure 4.6 depicts the strategy used to produce the pFS-BBL-monomer and the pFS-BBL-dimer construct.

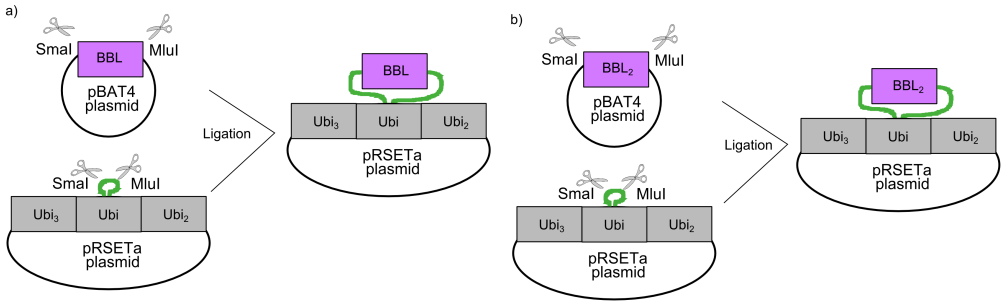

Figure 4.6: The schematic shows the approach used to build the pFS BBL-monomer a) and the pFS BBL-dimer b) construct.

To produce the pFS-BBL-monomer construct, both the before ordered BBL sequence inside pBAT4 (Clonetech) and the pFS vector were cut with the corresponding digestion enzymes SmaI and MluI and ligated together. For the pFS-BBL-dimer construct, the BBL-dimer was produced inside the ordered pBAT4 vector using the digestion enzymes BamHI and BglII. In this way the BBL-dimer could be cut with the digestion enzymes SmaI and MluI afterwards and also ligated into the pFS vector. The ligation step for both constructs was followed by a 2%-DNA agarose gel in order to confirm the correct insertion. Finally the two constructs were then transformed into the C41 competent cells in order to start the protein expression.

For purification the general purification protocol (see chapter 4.1.2) was followed. However for the second HPLC size exclusion superdex run the washing buffer was changed to the following content using 1.5M GdnHCl denaturant. Here the purification protocol was followed from [174].

Buffer	Superdex elution with denaturant
Buffer pH 7.5 Tris	50mM
NaCl	500mM
GdnHCl	1.5M
TCEP	2mM

Table 4.6: Washing buffer content with denaturant used for the pFS purification during the second HPLC run with the superdex size exclusion column

Afterwards the denaturant was removed against the measuring buffer 1xPBS pH 7.4 by using also ultracentrifugation Amicon 3k filters. Then a protein SDS gel was prepared to confirm the expression of the two constructs. Here the amount of purified protein was not as much as for the polyprotein constructs before (≈2mg).

Summary

All the produced polyprotein contructs are listed in figure 4.7.

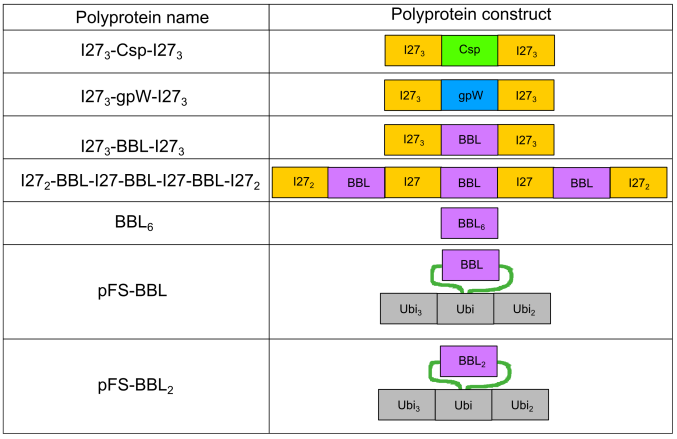

Figure 4.7: Overview of the built polyprotein constructs used throughout the thesis.

Table 4.7 gives an overview of the oligonucleotides used for PCR or sequencing.

Oligo name	Sequence 5'-3'	bp
BamHI-I27(start)	GAGAGA-GGATCC-CTAATAGAAGTGGAAAAGC	31
I27(end)-BglII-[C-C]-stop-HindIII	GAGAGA-AAGCTT-CTA-[-GCA-GCA-]-AGATCT-CCATTCTTTCACTTTCAG	45
Primer 1 Csp fwd	GAATTGAGATCCCGTGGCAAAGT	23
Primer 2 Csp rev	GTGGTGGAAAGATCCCTAATAGA	23
Primer 1 gpW fwd	GAATTGAGATCCGGCGGCGGTGG	23
Primer 2 gpW rev	GGCCCGGCGGGCAGATCCCTAAT	23
Primer 1 BBL fwd	GAAAGTGAAAGAATTGAGATCCC	23
Primer 2 BBL rev	GCAGCAAGATCTCTTTTTCGCCG	23

Table 4.7: Overview of the ordered oligonucleotides

4.2 Single Molecule Force Spectroscopy (SMFS) experiments

All the SMFS experiments were done by following protocols developed in the Prof. Julio Fernandez ([115] or see http://fernandezlab.biology.columbia.edu/) and Dr. Mariano Carrion lab([194] or see http://carrionvazquez-lab.org) with minor changes.

The two AFM instruments used were the AFM MultiMode equipped with a PicoForce module (Bruker/Veeco, www.bruker.com) and the force clamp AFM (Luigs-Neumann http://www.luigs-neumann.com) used in the Dr. Raul Jimenez (http://www.nanogune.eu/). Both AFMs can be equally used to apply a mechanical force on the protein of investigation. However there exist important differences between both instruments. The AFM MultiMode setup is mainly designed to perform imaging measurements with a high resolution accuracy reaching the sub nm range. However equipped with the PicoForce module and the PF-AFM-scanner, SMFS experiments are possible in the constant velocity mode [195].

In comparison, the force clamp AFM is especially designed to perform SMFS measurements. Here the implemented feedbackloop controls the mechanical force applied to the polyprotein sample during the experiments. This enables experiments under a constant force or under a controlled change in force with time (force ramp) additional to the constant velocity measurements[115].

The resolution in displacement of both instrument depends mostly on the acoustical isolation against vibrations, on the type of cantilever and finally on the experimental conditions

used. For the MultiMode AFM the force resolution was around ± 9pN standard deviation at 400nm/s in constant velocity measurements. For the Force Clamp AFM the force resolution was around ± 5pN standard deviation at 400nm/s for the same cantilever to have a comparison in constant velocity measurements (MLCT cantilever; spring constant k ∼ 10-20pN/nm). Within force clamp measurements the force could be hold constant depending of the cantilever used in a range of ± 2-3pN standard deviation at 20pN (BioLever; spring constant k ∼ 5pN/nm). The corresponding length displacement was ± 1nm standard deviation at 20pN (BioLever; spring constant k ∼ 5pN/nm). The force feedback time depends stronger on the explicit experiment and on the type of cantilever but in general was around 1-5ms for resetting the force after the unfolding of a titin I27 domain.

4.2.1 Experimental setup

A SMFS experiment was mainly set up in three steps, which are visualized in figure 4.8.

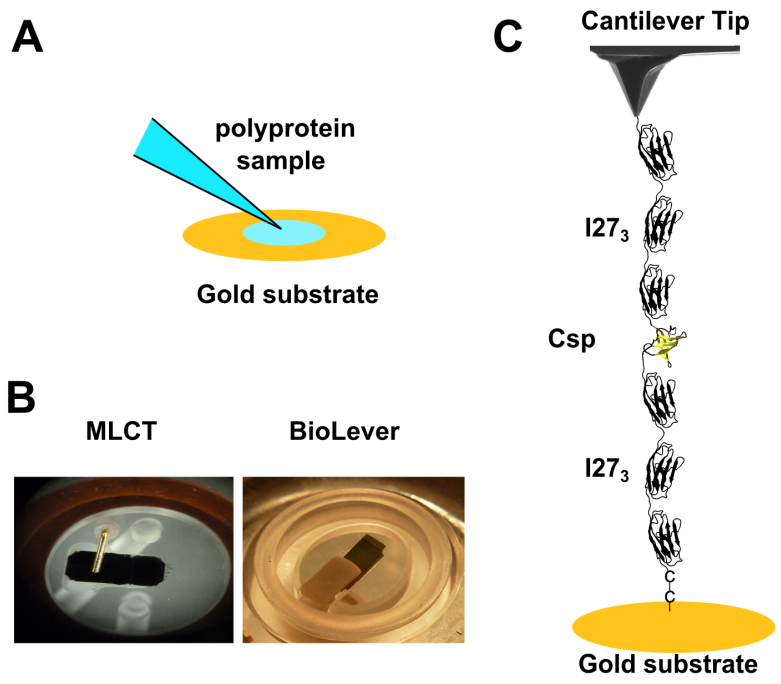

Figure 4.8: Conduction of a SMFS experiment. A) The protein sample is placed on the gold substrate. B) The cantilever is mounted into the fluid cell of the AFM. Here the two cantilever models MLCT and BioLever used in this work are shown. C) The AFM cantilever stretches the polyprotein construct, e.g. here the $I27_3$-Csp-$I27_3$ is shown including the two cysteine residues (C) at the end.

During the first step (figure 4.8 A), normally 40μl of freshly purified polyprotein sample (10-20 dilution with 1xPBS + 1mM TCEP; final concentration ≈ 0.05-0.1mg/ml) was pipetted on a commercially purchased evaporated gold substrate (Arrandee, http://www.arrandee.com/). The protein solution was left for around 20min on the sub-

strate for adsorption and then washed twice with 1xPBS buffer to remove unbound protein sample.

In the second step (figure 4.8 B) the cantilever was placed into the corresponding AFM fluid cell after UV-light sterilization. AFM cantilevers are typically manufactured from silicon nitride and they mostly have a gold coating on the back in order to improve their reflectivity. On the other side, which faces towards the sample the cantilever has a tip with a sharpness in the range of 10-30nm. The majority of the experiments were conducted with the MLCT cantilever (Bruker). Here either the cantilever C (spring constant k $\sim$ 30-40pN/nm; resonance frequence $f_R \sim$ 3-4kHz in buffer) or D (spring constant k $\sim$ 10-20pN/nm; resonance frequence $f_R \sim$ 1kHz in buffer) was used. Part of the experiments were also conducted with the BioLever cantilever (Olympus) with either a spring constant k $\sim$ 30-35 pN/nm (cantilever A, resonance frequence $f_R \sim$ 6-7kHz in buffer) or k $\sim$ 3-6pN/nm (cantilever B, resonance frequence $f_R \sim$ 1-2 kHz in buffer). The AFM fluid cell was then filled with the measuring buffer 1xPBS pH 7.4 in order to perform measurements in quasi-physiological conditions.

In the third step figure (4.8 C) the fluid cell was mounted into the AFM head. Then the slope of the cantilever deflection when being in contact with the gold surface was estimated and the spring constant was calibrated both with a build in function inside the AFM software using the thermal fluctuations method, which is based on the equipartition principle (see chapter 2.2.2 Force calibration) for both AFM setups. Therein first the thermal fluctuations of the cantilever far from surface are measured and fourier transformed to estimate the Power Density Spectrum (PDS) of the cantilever. The spring constant is then measured in integrating the area below the first resonance peak. Then in a second step the slope of deflection vs. extension in the contact region of the cantilever and the surface is being estimated. Afterwards the the SMFS experiment could be started.

4.2.2 Conducted SMFS measurements

$I27_3$-BBL-$I27_3$ construct

This BBL sample was measured with the Multimode AFM at a constant velocities of 400, 1700nm/s and 3500nm/s with using the MLCT cantilever (cantilever C, k $\sim$ 30-40pN/nm, $f_R \sim$ 3-4kHz in buffer).

$I27_2$-BBL-I27-BBL-I27-BBL-$I27_2$ construct

This BBL sample was measured with the Multimode AFM at a constant velocities of 400 and 1700nm/s with using the MLCT cantilever (cantilever C, k $\sim$ 30-40pN/nm, $f_R \sim$ 3-4kHz in buffer). All experiments were conducted in 1xPBS buffer (pH 7.4).

pFS-BBL construct

This BBL sample was measured with the Multimode AFM at a constant velocities of 400 and 1700nm/s with using the MLCT cantilever (cantilever C, k $\sim$ 30-40pN/nm, $f_R \sim$ 3-4kHz in buffer). All experiments were conducted in 1xPBS buffer (pH 7.4).

$pFS - BBL_2$ construct

This BBL sample was measured with the Multimode AFM at a constant velocities of 1700nm/s with using the MLCT cantilever (cantilever C, k $\sim$ 30-40pN/nm, $f_R \sim$ 3-4kHz in buffer). All experiments were conducted in 1xPBS buffer (pH 7.4).

BBL_6 construct

This BBL sample was measured with the Multimode AFM at a constant velocities of 400nm/s with using the MLCT cantilever (cantilever C, k $\sim$ 30-40pN/nm, $f_R \sim$ 3-4kHz in buffer). All experiments were conducted in 1xPBS buffer (pH 7.4).

$I27_3$-Csp-$I27_3$ construct

The Csp sample was measured with the Multimode AFM at a constant velocity of 10 and 400nm/s. Here the MLCT cantilever was used (cantilever C, k$\sim$30-40pN/nm, $f_R \sim$ 3-4kHz in buffer). Furthermore force ramp experiments at rates of 20, 80, 200 and 800pN/s and force clamp measurements at force of 20, 40, 60 and 80pN using the the Force-Clamp AFM were done. Herein additional experiments were conducted like using a constant force 20pN for up to 6 minutes and force clamp experiments with a sequence of forces from 20 to 80pN in one trace. For the Force-Clamp measurements the MLCT cantilever was used (cantilever D, k$\sim$10-20pN/nm, $f_R \sim$ 1kHz in buffer). All experiments were conducted in 1xPBS buffer with pH 7.4.

$I27_3$-gpW-$I27_3$ construct

The gpW sample was measured with the Multimode AFM at a constant velocity of 400nm/s using the MLCT cantilever (cantilever C, k $\sim$ 30-40pN/nm, $f_R \sim$ 3-4kHz in

buffer). Force ramp experiments at rate of 1, 20, 40, 60 and 80pN/s were conducted using the BioLever cantilever (cantilever B, k $\sim$ 3-6pN/nm, $f_R \sim$ 1-2 kHz in buffer) with the Force-Clamp AFM setup. Furthermore additional experiments were done like a combined force-ramp and force clamp experiment. Therein a force ramp of 1pN/s were set until 5pN were reached. Then the force was hold constant at 5pN for 30-40s, whereafter again the force ramp of 1pN/s was set to unfold to protein sample. Also these measurements were done with the BioLever cantilever (cantilever B, k $\sim$ 3-6pN/nm, $f_R \sim$ 1-2 kHz in buffer). All experiments were conducted in 1xPBS buffer with pH 7.4.

4.3 Analysis methods and software

4.3.1 Analysis of SMFS experimental data

All experimental data was analyzed with the IgorPro software (www.wavemetrix.com) as it is implemented within the force clamp AFM. The analysis procedure file was a kindly gift from the Prof. Julio Fernandez lab (AFM_Analysis_V2.40.ipf; see http://fernandezlab.biology.columbia.edu/). Additionally, in order to be able to also analyze the constant velocity data achieved with the MultiMode AFM, a procedure file was written to import the data into IgorPro and analyze it with the common analysis procedure file.

All force extension traces from the constant velocity measurements were fitted to the discussed WLC model (see section 2.2.2 and formula 2.2) to estimate the difference in contour length ΔL. Normally the last peak of the sawtooth pattern was fitted to a certain persistence length p value between 0.3 and 0.5nm, which was used to fit all the peaks in the entire construct.

To measure the lifetimes of the constant force measurements at 5pN for the gpW protein, a already build procedure (Name: Detect steps ; Function: FC_DetectSteps) inside the analysis software (AFM_Analysis_V2.40.ipf) of the Force-clamp AFM has been used. This written function calculates the average values before and after a given point of the trace. The number of point used to calculate the average can be changed with the Slide Size value. If the difference between the calculated average value before and after the point exceeds a given threshold (Min stepsize), the point will be saved and its given time and length value. From the distribution of point the lifetimes in the corresponding state can be estimated through subtraction. To conduct the DetectSteps function the traces were smoothed using the Median 10 option, which calculates the average value of the 10 points

before and after each point of the trace. This was necessary in order to find later the points at which the length values changes from on state to another. With using a threshold value of 3nm and a slide value of 1 the procedure function was applied on the hopping region of the trace. The correctness of the procedure in finding the relevant points was then manually confirmed and if needed not founded points were added or wrongly estimated points were deleted.

In some cases the bootstrap method has been used and will be mentioned in the corresponding section. One of its applications is to estimate the mean and its standard deviation of an unknown distribution. Here the bootstrap method was used to estimate the standard deviation from the average times of the dwell time distribution measured in constant force AFM experiments. The bootstrap method relies on random sampling with replacement of the input data. Each of the new bootstrap sample has a changed mean value depending on the number n of replacement. Therefore depending on the number of iterations a distribution of the mean values of the dwell time distribution can be obtained and its standard deviation can be estimated, although the original distribution is e.g. left shifted. This method was applied in using a statistical package within the IgorPro software (StatsResample).

4.3.2 K-means cluster analysis

The K-means algorithm is a common used method for partitioning n multidimensional data points into K clusters $C = \{c_k, k = 1, ..., K\}$. Here the number of clusters K is initially determined and the cluster algorithm is performed until the given number K fits the overall data (n multidimensional data) best [196]. The purpose of the algorithm is to divide the data into K partitions that the sum of the squared deviations from the cluster centers (centroids) are minimal. This algorithm can be formulated with the following function:

$$J(C) = \sum_{k=1}^{K} \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \tag{4.1}$$

with $\|x_i - \mu_k\|^2$ describing the squared error between the data point x_i and the cluster center μ_k for C_k clusters. The K-means clustering algorithm starts then with an initial guess of the number of K clusters and assigns then patterns in order to reduce the squared error.

Here the in-written Matlab (www.mathworks.com) K-means algorithm has been used to conduct the K-means cluster analysis of the input data from the constant force SMFS

experiments on the Csp for 20000 iterations, each starting with a new set of initial cluster centroids positions. The output was then the mean of the centroids with the smallest data-point-to-centroid distance and its corresponding standard deviation.

4.3.3 Stochastic kinetic simulation of SMFS force ramp and constant force experimental trajectories

The stochastic kinetic simulation was implemented using a procedure similar to the original Gillespie stochastc simulation algorithm (SSA) [197, 198].

In particular, when assuming a two state process for simplicity, the protein can only be in the folded or the unfolded state depending on its given microscopic rates k_f^0 and k_u^0. Under a mechanical force both rates change, a process which was approximated with the Bell equation $k_f(F) = k_f^0 e^{-\frac{F\Delta x_f}{k_B T}}, k_u(F) = k_u^0 e^{\frac{F\Delta x_u}{k_B T}}$ where $k_{f/u}^0$ is the folding/unfolding rate constant in the absence of force, F is the stretching force, $\Delta x_{f/u}$ is the distance to the folded and unfolded transition state, k_B is the Boltzmann constant and T is the absolute temperature.

Furthermore the length of the folded state was estimated with the experimentally determined transition to the unfolded extended state (Δx). The length of the extension x(F) was estimated by solving the WLC model numerically for dF/dx with $F(x) = \frac{k_B T}{p}\left[\frac{1}{4}\left(1-\frac{x}{L_C}\right)^2 - \frac{1}{4} + \frac{x}{L_C}\right]$ and L_c being the contour length of the protein and ρ being the persistence length. The stochastic kinetic simulation similar to that of Gillespie's was now the following:

For a given timestep Δt the protein can either stay in its state or jump to the other state as defined by the folding or unfolding rate at the applied force. A stochastic time series is now produced by drawing time steps δt from an exponential distribution $\exp(1/k_f)$ or $\exp(1/k_u)$ depending on the state in which the protein is at that time step. If the randomly drawn time step δt is smaller than the time step of the iteration Δt, which correspond to the sampling rate of the simulation, than the protein jumps to the other state.

Additionally the thermal fluctuations due to Brownian motion of the cantilever with a spring constant k_c were implemented into the simulation assuming that the free cantilever oscillation is small and can be described as that of an harmonic oscillator. Then when using the potential of an harmonic oscillator $U = 1/2k_c(x-x_0)^2$ its width can be estimated by the standard deviation σ. This results in $\sigma = \frac{k_B T}{k_c}$, which means that a smaller spring constant is causing larger length fluctuations. The force fluctuations are connected to the length fluctuations with Hooke's law for small displacements. Hence the spring constant

k_c effects the fluctuations in length and force caused by the cantilever itself during its free oscillation.

4.4 Steered Molecular Dynamics (SMD) Simulations

SMD simulation can provide the mechanical unfolding reaction of a protein with atomistic resolution, which is not accessible during the SMFS experiments. Especially the unraveling of structural elements during the application of a mechanical force on the protein structure can provide additional insights, which could help to interpret the experimental data. The major drawback of these simulations is that they have to be conducted with a higher force or constant velocity than under experimental conditions in order to be able to calculate the simulation in an affordable calculation time.

Here, SMD simulations were conducted with the three proteins of investigation Csp, BBL and gpW with the NAMD software and using the CHARMM22 forcefield [199, 200] All simulations were visualized using the VMD software [201]. For realizing the simulations the standard protocols, software scripts, simulation files (e.g. the topology and parameter file) and tutorials available from the Klaus Schulten group (http://www.ks.uiuc.edu/) have been used and followed for all conducted SMD simulations.

4.4.1 Setting up a SMD simulation

For all the three proteins the corresponding pdb files were downloaded and/or modified to adopt to the real structure used within the experiments and will be described in the next section and in the corresponding results and discussion section. Figure 4.9 gives an overview of the three steps followed to set up all SMD simulations.

During the first step the pdb file was downloaded and the structure file (psf) was generated with the VMD program plugin, using the topology file all27_prot_lipid.inp (figure 4.9 1a)). Then the vector connecting the N-C terminus of the structure was aligned parallel to the x-direction, which serves as the pulling coordinate (figure 4.9 1b)). Afterwards a waterbox was created around the protein, which dimensions depend on the size of the protein but also on its total unfolding length especially in the x-direction (figure 4.9 1c)). TIP3W water molecules were used for all the SMD simulations.

In the second step, the NAMD simulation file was generated using a VMD software plugin and the parm_all27_prot_na.prm file as the parameter file (figure 4.9 2)). Furthermore for all simulations periodic boundary conditions (PBC), constant NPT conditions (constant number of atoms N, constant pressure P and temperature T) and langevin dynamics

(constant temperature) were used. All simulations were calculated with a timestep of 1fs. This NAMD file was written to conduct the minimization and equilibration MD simulation. Therein typically the system was minimized using deepest decent for 2000 timesteps and thereafter equilibrated for 1ns using NAMD. One example for such a typical namd.file used is shown in the appendix (see Apendix 1 *V*).

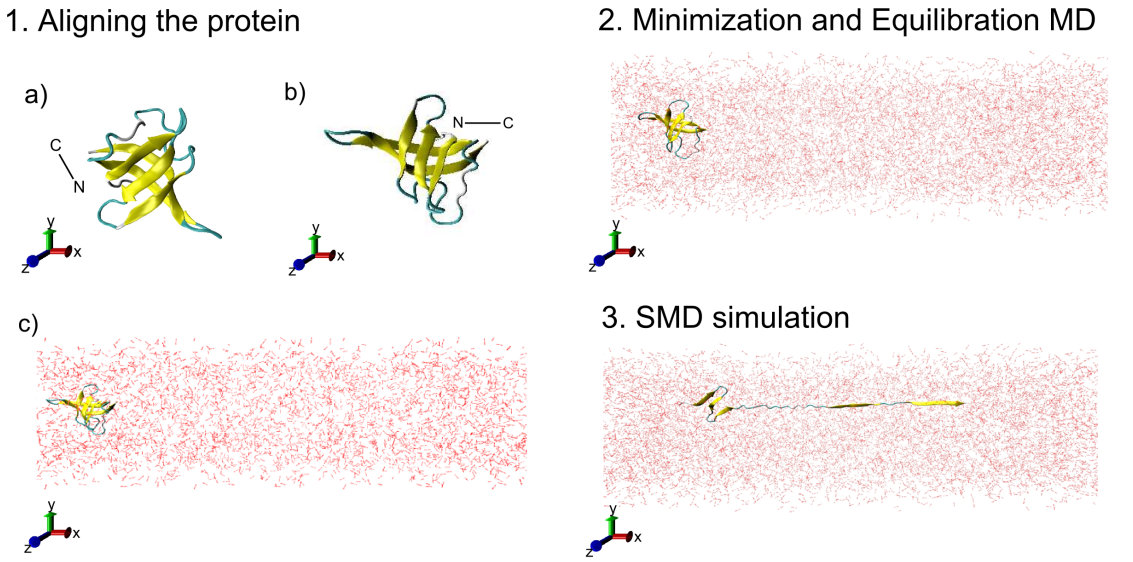

Figure 4.9: The procedure used to set up an SMD simulation is shown. 1) Aligning the pdb file; a)downloading the pdb file, b) aligning the structure and c) constructing the waterbox 2) Minimization and enquilibration MD 3) SMD simulation

Then during the third and last step, SMD simulations were performed using the coordinate file from the equilibration MD simulation (figure 4.9 2)). Usually the C_α-atom of the N-terminus of the protein was hold fixed, whereas the force was applied on the C_α-atom of the C-terminus of the protein being pulled in the x-direction. SMD protocols were used, which can apply a mechanical force with a constant velocity and a constant force. Both modes were conducted with each of the three proteins.

Each SMD trajectories were analyzed by measuring the force applied on the C_α-atom of the C-terminus vs its x-coordinate for constant velocity simulations and by measuring the distance between the C_α-atoms of the N and C-terminus vs time for constant force simulations. Furthermore the timeline plugin of the VMD software was used to track when secondary structure elements get disrupted with time during an SMD simulation.

4.4.2 Conducted SMD simulations

BBL

In the case of the BBL two different files were used for SMD simulations on BBL. First the short fragment 2CYU.pdb file ([83]) was used for SMD simulations. Here one constant velocity SMD simulations were conducted at a speed of 1m/s.
However as the 2CYU.pdb BBL structure did not represent the structure used during the experiments, the 2WXC.pdb ([202]) file was downloaded and modified to adopt the used BBL structure used within the experiments (see figure 3.2). The modifications affected both tails. In the N-terminus tail the three residues glycine, serine and glutamine (GSQ) were replaced by two lysines (KK). To the C-terminus tail the four residues proline, alanine and two lycines (PAKK) were added. The tryptophan W18 was changed to histidine H18, as it is present in the experimentally used structure. With this longer structure one constant velocity SMD simulations was conducted at 1m/s for comparison. Here additionally, one constant force SMD simulation was done at 50pN, 125pN, 150pN, 200pN, 300pN and 500pN. Three constant force SMD simulations wee conducted at 100pN.

Csp

To simulate the Csp, the 1G6P.pdb file ([184] was downloaded and used for all SMD simulations (see figure 3.3).
Five constant velocity SMD simulations were conducted with a velocity of 1m/s. 17 constant force SMD simulationes were performed at 200pN, one at 300pN and one at 400pN. During four from the total of 17 constant force simulations at 200pN, the protein was also pulled in the opposite direction. I.e the C_α-atom of the C-terminus of the protein was kept fixed and the force was applied on the C_α-atom of the N-terminus of the protein being pulled in the x-direction.

gpW

For the gpW protein, the 2L6Q.pdb file ([190] was downloaded (see figure 3.4). In order to adopt to the experimentally used sequence, three more residues (proline, alanine and glycine; (PAG)) were added to the C-terminus and five glycine residues (GGGGG) were added to the N-terminus in order represent the experimentally used structure.
One constant velocity SMD simulation was conducted at a velocity of 1m/s. One constant force SMD simulations was performed at 100pN, five at 150pN and one at 200pN.

Part III

Results and Discussion

Chapter 5

Mechanical unfolding of the BBL domain

This chapter reveals all the results describing the detected mechanical unfolding behavior of the BBL domain, which will be divided into the following sections. The first section will discuss all conducted constant velocity SMFS experiments on the five designed BBL containing polyprotein constructs. These constructs are the $I27_3$-BBL-$I27_3$, the $I27_2$-BBL-I27-BBL-I27-BBL-$I27_2$, the pFS-BBL, the pFS-BBL-Dimer and the poly-BBL_6 construct (see also figure 4.7 for an overview). Then the results of the SMD simulations on the single BBL structure will be given.

5.1 SMFS measurements on the BBL domain

In all constant velocity measurements at the beginning of the trace the cantilever was pushed first with $|\tilde{F}| = 1$nN against the surface, before the cantilever is retracted. With that amount of force protein adsorption to the cantilever was ensured. In each force/extension trajectory the red trace represents the approach and the blue trace the retraction of the cantilever from the surface. The insets in each trajectory figure show the unfolding signal of the BBL a higher magnification.

Normally not every approach/retraction cycle of the cantilever to the surface results in a adsorption of the polyprotein sample to it. Furthermore around 90% of traces which indicate protein adsorption need to be discarded, which is illustrated in figure 5.1. Therein SMFS measurements at a constant velocity of 400nm/s on the $I27_3$-BBL-$I27_3$ construct are shown. As BBL is expected to have a lower unfolding force than the titin I27 domain as it consists only of three small α-helices, its unfolding signal should appears always at

the beginning of the trace.

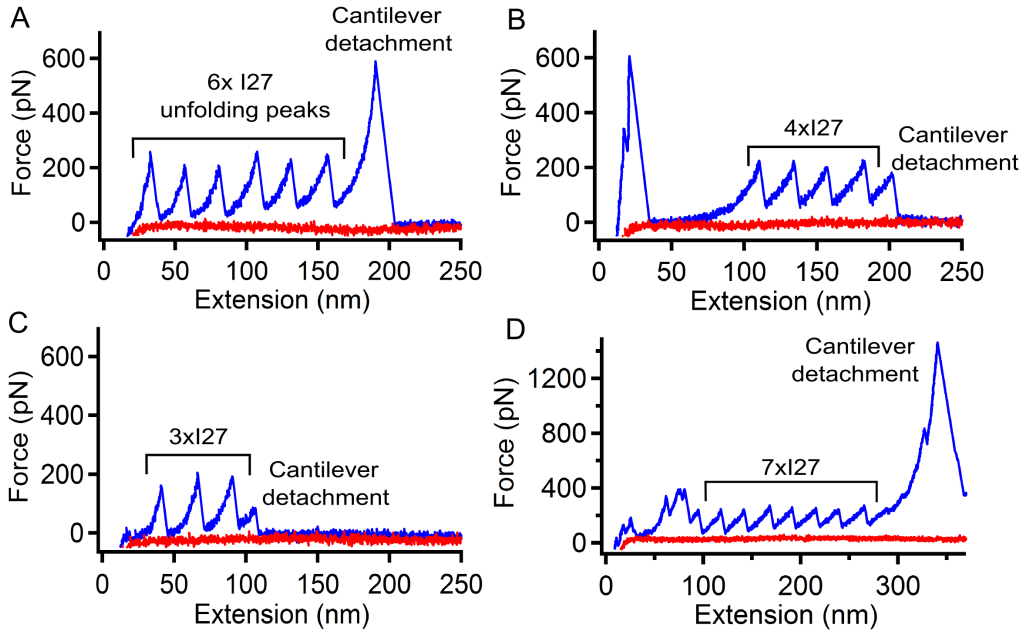

Figure 5.1: Four force extension traces (A)-(D) for BBL inside the $I27_3$-BBL-$I27_3$ cosntruct at a constant velocity of 400nm/s, as examples which cannot be used for the data analysis. The I27 signal and the cantilever detachment peak are indicated in each trace.

The four examples shown in figure 5.1 (A)-(D) have to be singled out from the traces used for the data analysis for the following reasons:

The first trace (figure 5.1 A) shows an example when the first titin I27 domain unfolds very close to the surface. Although all six I27 titin domains unfold, the beginning of the trace before the first I27 unfolding peak cannot be fitted to the WLC model.

In the second trace (figure 5.1 B) the unfolding of the titin I27 domains occur too late in the trace (at around 90-100nm). This huge distance already indicates that there is some more adsorption (see noise at the beginning of the trace) to the cantilever than just one single $I27_3 - BBL - I27_3$ polyprotein construct. Although four I27 unfolding peaks are detected these kind of trace were discarded. The noise at the beginning of the trace is often caused by interactions of the cantilever and the surface during the approach/retraction cycle.

The third example (figure 5.1 panel C) shows a clean trace but just three I27 titin domains are unfolded, before the cantilever gets detached from the polyprotein sample. In this case when detecting only three I27 titin unfolding peaks it cannot be made sure that the mechanical force was also applied to the BBL protein as within the $I27_3 - BBL - I27_3$ polyprotein construct the BBL domain is flanked by three I27 domains on each side. Therefore also these kind of traces were not selected for the data analysis.

Finally, the last example (figure 5.1 D) shows multiple adsorption on the cantilever from at least two polyproteins, because seven titin I27 domains unfolding peaks are detected and also the beginning of the traces shows much noise. Although in this case the BBL unfolding signal could be present at the beginning of the trace, these kind of trace were discarded from the data pool used for the analysis.

Thus all these examples show frequently occurring traces, which cannot be used for the data analysis process when measuring the I27 titin domain containing polyprotein constructs. Hence good force extension traces need to be selected, from which examples will be shown in the following two sections. This serves as a general guide of trace selection for good and usable force extension curves followed throughout this work.

As it has been explained in figure 5.1 (A) and (B), the complete unfolding length of the polyprotein sample has to be in a certain range before the first titin I27 domain unfolds, which is constrained by the sum of the unfolding length of the BBL (around 16-20nm) and the length of the six folded titin I27 domains (26.4nm) and lies in this case between 16-20nm and 42.4-46.4nm. Here the total length of BBL was roughly estimated by multiplying its number of amino acids 50 (40aa without tails) by the approximated length of a single amino acid, which is 0.4nm. The length of the six folded I27 domains was estimated by multiplying it folded length of around 4.4nm by 6, which was estimated from its pdb file 1tit.pdb. In the case of the $I27_2$-BBL-I27-BBL-I27-BBL-$I27_2$ construct this range lies between 48-60 and 74.4-86.4nm, because this constructs contains three BBL domains.

In the case of the pFS constructs containing the BBL and the BBL-dimer, trajectories were collected in which the characteristic length of the ubiquitin carrier was present. The big advantage of using this type of construct is that the BBL always unfolds after the unfolding of the ubiquitin carrier, therefore its signal is shifted away from the beginning of the force extension trace. Therefore less traces need to be discarded, because basically just the unfolding peak of the ubiquitin carrier needs to be present in the force extension curve in order to clearly identify the signal of the BBL.

In contrast, a molecular mechanical fingerprint is absent in the poly-BBL_6 construct, here only its total unfolding length was used as a reference to select force extension traces.

For the data analysis, the unfolding force at each peak was directly measured and the difference in contour length (ΔL) was estimated by fitting each peak to the WLC model (see WLC equation 2.2 in chapter 2.2.2). The measured contour lengths (ΔL) of the BBL, titin I27 or ubiquitin will be indicated in the figures.

5.1.1 Constant velocity measurements on the $I27_3$-BBL-$I27_3$ construct

The $I27_3$-BBL-$I27_3$ construct was measured at a constant velocity of 400, 1700 and 3500nm/s using the MultiMode AFM.

Constant velocity measurements at 400nm/s

Figure 5.2 shows four typical force extension traces measured at a constant velocity of 400nm/s.

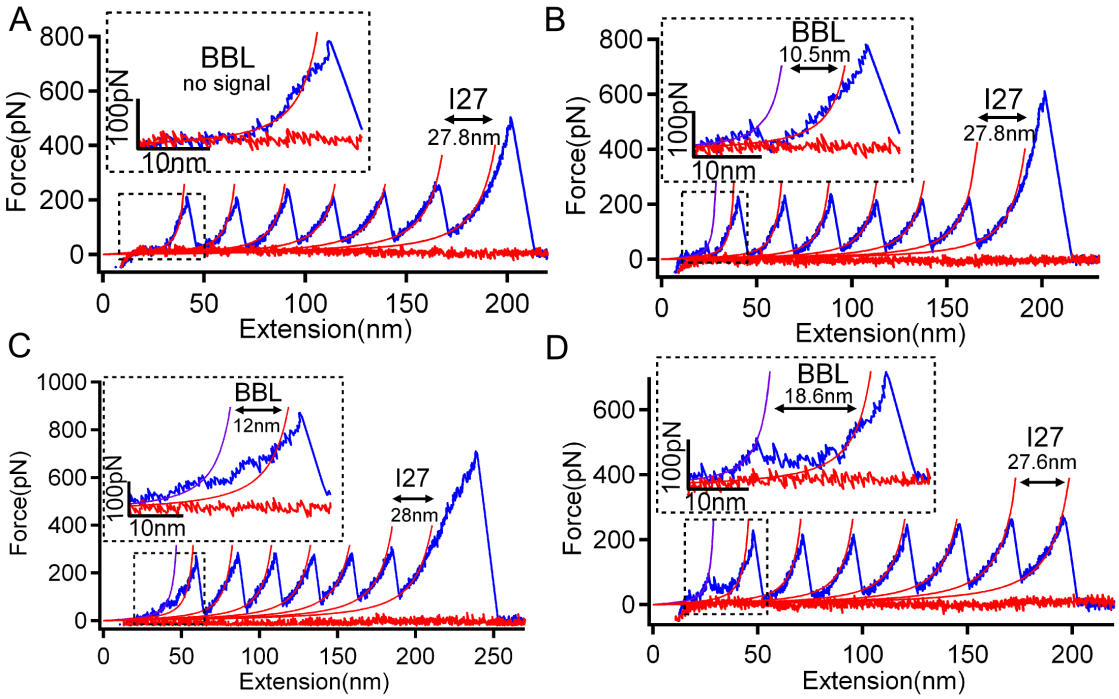

Figure 5.2: Force extension traces for BBL at a constant velocity of 400nm/s (A) No detected BBL unfolding signal (B) BBL peak unfolding signal (C) BBL hump unfolding signal (D) complex BBL unfolding signal

All four depicted traces show the complete $I27_3$-BBL-$I27_3$ unfolding, therefore all the six titin I27 unfolding peaks can be seen. The unfolding pattern of BBL was found to be very heterogeneous within the measured and selected traces (n=101). This heterogeneity was clustered into the three groups. The first group contain traces, in which no BBL unfolding signal could be detected, although the corresponding force extension trace was identified to be a clean and good trace and therefore following the conditions mentioned before (see figure 5.1). Around 25% of all traces (n=25) did not contain any distinguishable BBL signal (see figure 5.1 A).

Force extension traces, which revealed a unfolding peak of the BBL were put in the second group. Here around 43% of all traces (n=43) showed this shape of the mechanical unfolding

behavior of BBL (see figure 5.1 B).

Additionally traces were identified in which BBL unfolds more through a hump than via a single peak. The remaining 32% traces (n=33) were therefore put in the third group (see figure 5.1 C). In some traces it was difficult to distinguish between a single peak and a hump (see 5.1 D), which illustrates the high degree of heterogeneity in the unfolding behavior of the BBL domain.

Figure 5.3 gives an overview of the measured distributions of unfolding force and difference in contour length of the BBL and for comparison also of the titin I27 domain. In the case of the BBL, just the beginning of the unfolding signal was fitted to the WLC model to approximate the total unfolding length of the BBL (especially for the traces detected in panel D). Also the corresponding unfolding force was just measured at this point at the beginning of the BBL unfolding signal.

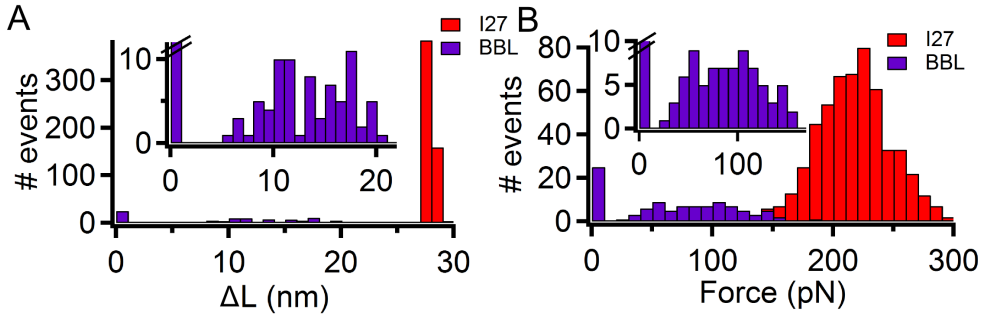

Figure 5.3: Histograms of the estimated difference in contour length (bin size 1nm) (A) and the measured unfolding force (bin size 10pN) (B) of the BBL protein and the titin I27 domain at a constant velocity of 400nm/s

Averaging the distributions for the BBL results in a mean unfolding force of (68.6 ± 49.9)pN and in a mean difference in contour length of (10.0 ± 6.6)nm. The average values contained also the counted 25 traces in which no BBL signal could be measured (zero force and zero length). The corresponding results from the average of the titin I27 domain are (218.5 ± 32.6)pN and (27.7 ± 0.4)nm, which coincides with the corresponding results for the I27 from the literature (e.g. see [167]).

Constant velocity measurements at 1700nm/s

Four typical force extension traces measured at a constant velocity of 1700nm/s are illustrated in figure 5.4. All four depicted traces show the complete unfolding of the $I27_3$-BBL-$I27_3$ construct. The unfolding behavior of the BBL domain was not found to be very different from the measurements done at 400nm/s, as the partitioning into the three selected groups was not changing notably. From all traces (n=115) around 28% (n=32)

showed no BBL unfolding signal (see figure 5.4 A). 40% (n=46) of all traces showed one unfolding peak for the BBL domain (figure 5.4 B). In the rest of 32% (n=37) traces the BBL protein was unfolding through a hump (figure 5.4 C). Figure 5.4 D shows again an example in which the classification of the BBL unfolding behavior into a group was more difficult.

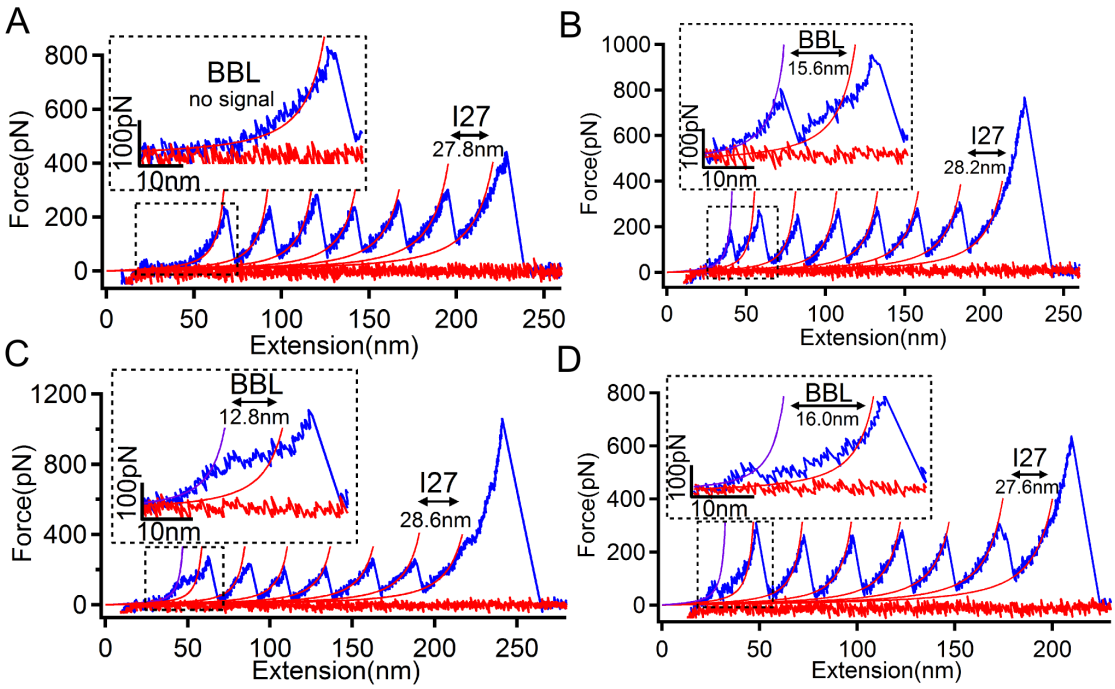

Figure 5.4: Force extension traces for BBL at a constant velocity of 1700nm/s (A) No detected BBL unfolding signal (B) BBL peak unfolding signal (C) BBL hump unfolding signal (D) complex BBL unfolding signal

Figure 5.5 gives an overview of the measured distributions of unfolding force and difference in contour length of the BBL domain and for comparison also of the titin I27 domain.

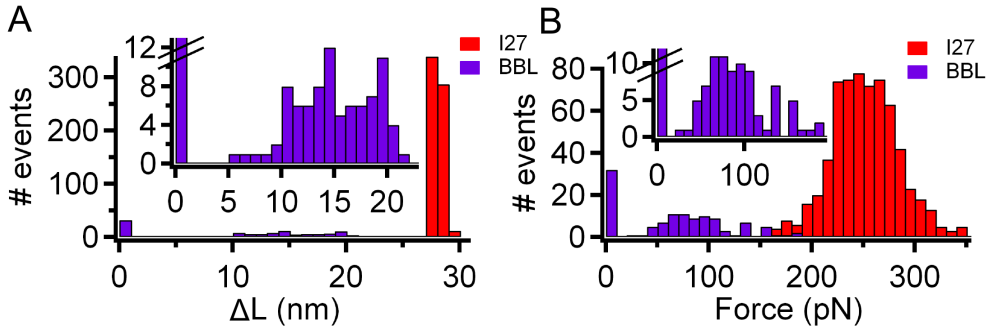

Figure 5.5: Histograms of the estimated difference in contour length (bin size 1nm) (A) and the measured unfolding force (bin size 10pN) (B) of the BBL protein and the titin I27 domain at a constant velocity of 1700nm/s

Here the averaging results for the BBL in a mean unfolding force of (69.5 ± 54.6)pN and in a mean difference in contour length of (10.9 ± 7.3)nm. The corresponding results from the average of the titin I27 domain are (251.5 ± 32.9)pN and (27.9 ± 0.3)nm.

Constant velocity measurements at 3500nm/s

At a velocity of 3500nm/s the unfolding behavior of the BBL domain was detected to be very similar to the measurements done at 400 and 1700nm/s, although the population of the three groups was changing a bit. Four typical force extension traces measured at a constant velocity of 3500nm/s are depicted in figure 5.6.

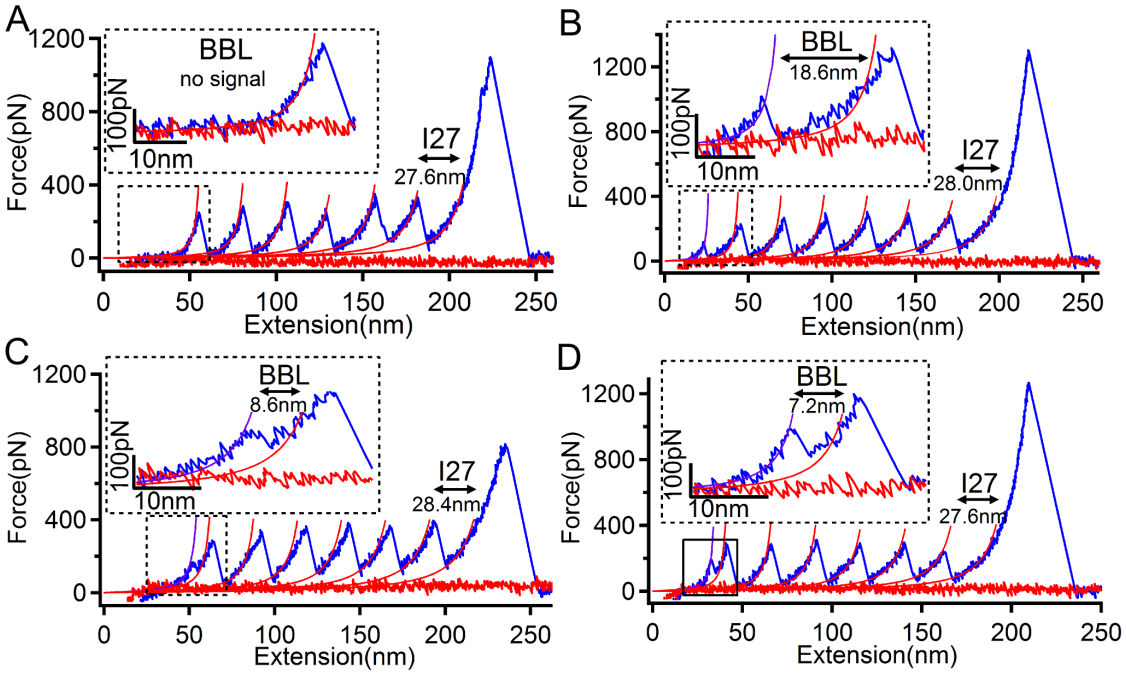

Figure 5.6: Force extension traces for the BBL domain at a constant velocity of 3500nm/s (A) No detected BBL unfolding signal (B) BBL peak unfolding signal (C) BBL hump unfolding signal (D) complex BBL unfolding signal

All four depicted traces show the complete unfolding of the $I27_3$-BBL-$I27_3$ construct, as all six titin I27 unfolding peaks are visible. From all traces (n=72) around 38% (n=27) showed no BBL unfolding signal (see figure 5.6 A), which is a slightly increased value compared to 25-30%, which was measured at lower velocities. 15% (n=11) of all traces showed one unfolding peak for the BBL (figure 5.6 B). In the majority of traces (47%) (n=34) the BBL was unfolding through a hump (figure 5.6 C). Also at this high velocity in some traces the unfolding behavior of the BBL domain revealed heterogeneous unfolding patterns (see 5.6 D).

The corresponding distributions of the unfolding force and difference in contour length of

the BBL domain of the titin I27 domain are depicted in figure 5.7.

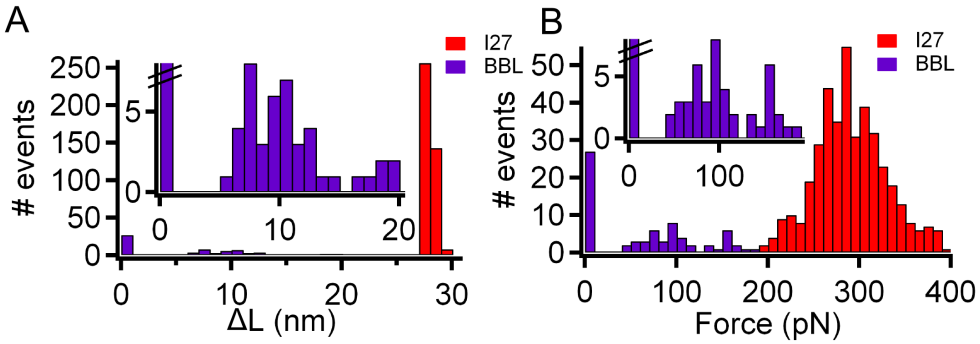

Figure 5.7: Histograms of the estimated difference in contour length (bin size 1nm) (A) and the measured unfolding force (bin size 10pN) (B) of the BBL protein and the titin I27 domain at a constant velocity of 3500nm/s

Averaging the distributions for the BBL results in a mean unfolding force of (64.5 ± 59.2)pN and in a mean difference in contour length ΔL of (6.6 ± 6.0)nm. The corresponding results from the average of the titin I27 domain are (289.7 ± 41.6)pN and (27.9 ± 0.3)nm.

Summary and Discussion

The constant velocity experiments done at 400, 1700 and 3500nm/s unveil a detectable mechanical response from the BBL domain. Furthermore the conducted measurements indicate a high complexity of the mechanical unfolding mechanism. The BBL unfolding curves could be partitioned into three different groups, however the force extension traces revealed a high heterogeneity of unfolding patterns. A high percentage of traces shows no BBL unfolding behavior (25-40%), which indicates that BBL unfolds in many cases below 15-20pN, which is in the range of the resolution limit of the used MultiMode AFM. The detected unfolding force and ΔL values for the I27 at the different velocities of the conducted measurements indicate a well calibrated instrument. The following table 5.1 summarizes the results of the BBL and of the titin I27 domain at the three used velocities. Therein it can be seen that the change in velocity does not have an observable effect on neither the average value of the difference in contour length nor on the measured unfolding force of the BBL domain. Furthermore both distributions keep their broadness, which is represented by the std value which is almost as high as the average value estimated at each velocity. The fact that at a speed of 3500nm/s the estimated unfolding force and length are smaller than at lower velocities can be explained by the decrease of resolution or the increase in experimental noise, which could cause that more traces are found where no

BBL unfolding signal can be detected. This shifts the measured average to smaller values.

	BBL		
Force(pN)	68.6 ± 49.9	69.5 ± 54.6	64.5 ± 59.2
ΔL(nm)	10.0 ± 6.6	10.9 ± 7.3	6.6 ± 6.0
velocity (nm/s)	400	1700	3500
	titin I27		
Force(pN)	218.5 ± 32.6	251.5 ± 32.9	289.7 ± 41.6
ΔL(nm)	27.7 ± 0.4	27.9 ± 0.3	27.4 ± 0.4
velocity (nm/s)	400	1700	3500

Table 5.1: Overview of the measured force and difference in contour length ΔL of BBL and of I27

Whereas the unfolding force of the I27 domain increases with increased velocity, which is explained by the effect of force on the (un)folding rate of the protein (see Bell equation 2.3) and especially here by the change of force in time.
The next chapter will reveal the results obtained from the approach of measuring multiple domains of BBL (in this case three BBL domains) inside a I27 domain containing polyprotein construct. The idea of designing the $I27_2$-BBL-I27-BBL-I27-BBL-$I27_2$ was to identify a repeatable pattern, which can be referred to the consecutive unfolding of a single BBL domain. It was expected that a change in the number of BBL-domains would clearly identify the BBL unfolding signal.

5.1.2 Constant velocity measurements on the $I27_2$-BBL-I27-BBL-I27-BBL-$I27_2$ construct

Experiments were conducted at a constant velocity of 400 and 1700nm/s using the MultiMode AFM.

Constant velocity measurements conducted at 400nm/s

In general, the unfolding pattern of three BBL domains inside one construct was found to be even much more complex and heterogeneous than that from the single BBL domain containing polyprotein construct. Figure 5.8 depicts three force extension traces measured at a constant velocity of 400nm/s. The shape of the BBL unfolding signal was again distributed into the same three groups, which has already been used to describe the single BBL domain unfolding behavior. Around 21% (n=8) of all traces (n=38) showed no unfolding signal from any of the three BBL domains (see figure 5.8 A).

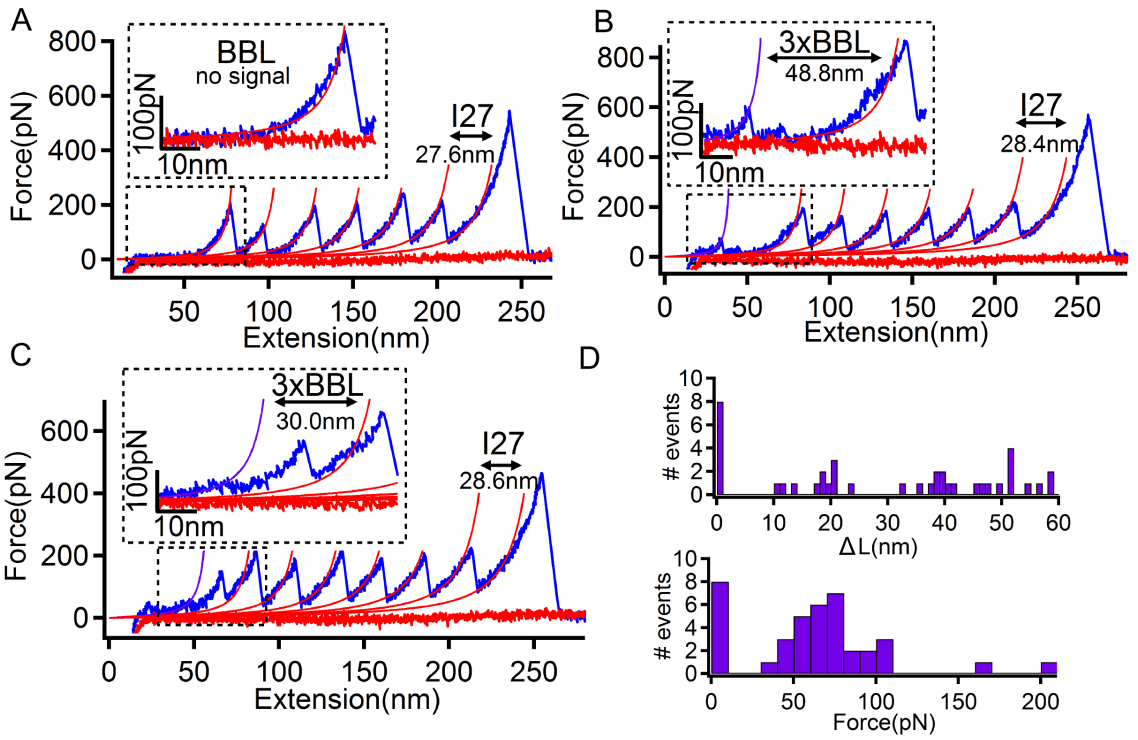

Figure 5.8: Force extension traces for the three BBL domain containing construct at a constant velocity of 400nm/s (A) No detected BBL unfolding signal (B) Several peaks detected for the unfolding behavior of the three BBLs (C) A mixture of hump and peak detected for the unfolding behavior of the three BBLs (D) Histograms of ΔL (bin size 1nm) and unfolding force (bin size 10pN) of the three BBL domains

The majority of traces 74% (n=28) showed an unfolding peak for either one or up to three BBL domains (figure 5.8 B). In the rest of 5% (n=2) traces the three BBL domains were discovered to unfold through a hump (figure 5.8 C). Interestingly, the unfolding signals of the three BBL domains could be detected with a difference in contour length of around 60nm before the first I27 titin domain unfolds. I.e. the three BBL domains are unfolding together or connected to each other, instead of a consecutive unfolding mechanism like e.g. the I27 domain does.

Figure 5.8 D gives an overview of the measured distributions of unfolding force and difference in contour length of the BBL. Here just the beginning of the unfolding signal was fitted to the WLC model to approximate the total unfolding length of the three BBL domains together, because no typical signal could be referred to a single BBL domain. Therefore the unfolding force was just measured at the point of fitting. Averaging the distributions for the three BBL results in a mean unfolding force of (67.5 ± 57.6)pN and in a mean value of ΔL of (28.1 ± 10.4)nm. The corresponding unfolding force of (200.0 ± 27.9)pN and difference in contour length of (28.0 ± 0.6)nm of the I27 domain was found to be in the same range as identified for the $I27_3$-BBL-$I27_3$.

Constant velocity measurements at 1700nm/s

Measurements at 1700nm/s also indicated a high complexity within the unfolding patterns of the three BBL domains together. Figure 5.9 shows three force extension traces measured at that velocity. All three depicted traces show six titin I27 unfolding peaks and therefore the complete unfolding of the $I27_2$-BBL-I27-BBL-I27-BBL-$I27_2$ construct. A difference to the experiments done at 400nm/s was not detected. Around 26% (n=8) of all traces (n=31) showed no BBL unfolding signal (see figure 5.9 A). An unfolding peak for at least one of the three BBLs was found in 44% (n=14) of all traces (figure 5.9 B). In the remaining percentage of 30% (n=9) of measured traces the BBL was discovered to unfold through a hump (figure 5.9 C).

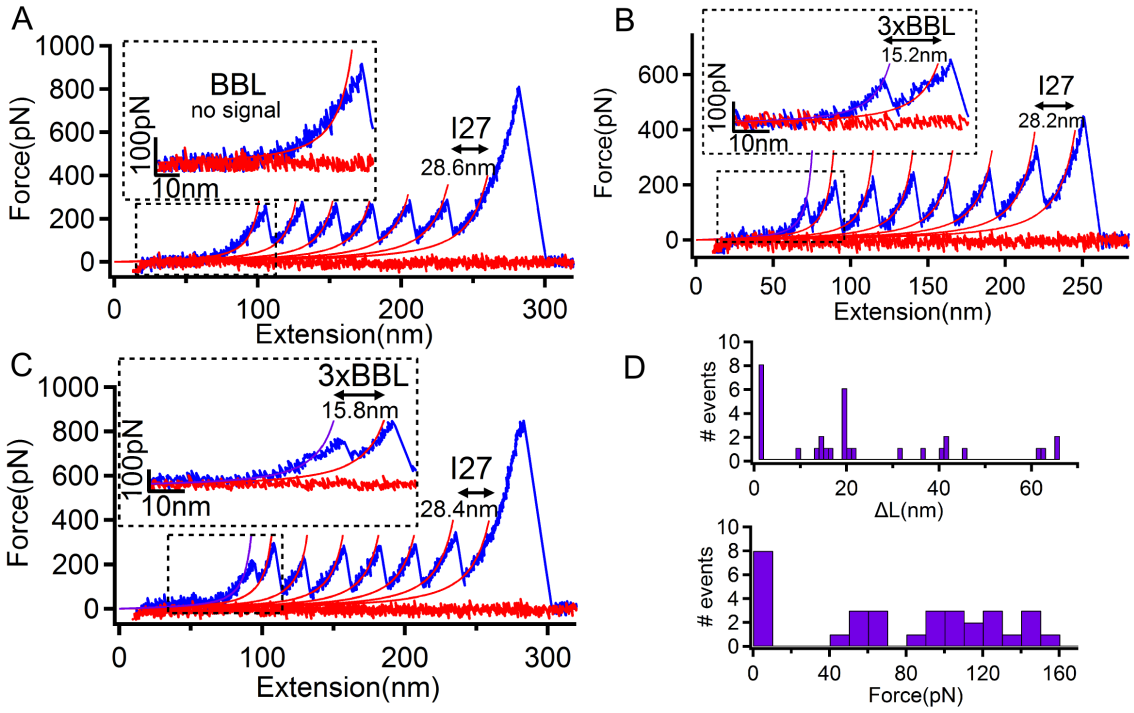

Figure 5.9: Force extension traces for the three BBL domains at a constant velocity of 1700nm/s (A) No detected BBL unfolding signal (B) Several peaks detected for the unfolding behavior of the three BBLs (C) A mixture of hump and peak detected for the unfolding behavior of the three BBLs (D) Histograms of ΔL (bin size 1nm) and unfolding force (bin size 10pN) of the three BBL domains

Panel D in figure 5.9 gives an overview of the measured distributions of unfolding force and difference in contour length of the three BBL domains together. The average for the three BBL gives a mean unfolding force of (74.8 ± 52.2)pN and a mean value of ΔL of (22.2 ± 20.3)nm. The corresponding unfolding force of (256.2 ± 31.6)pN and difference in contour length of (28.4 ± 0.5)nm of the I27 domain was found to be in the same range as identified with the $I27_3$-BBL-$I27_3$ at the same velocity.

Summary and Discussion

The conducted measurements show clearly that interpreting the data became even more complicated when increasing the amount of BBL domains in the polyprotein construct. Nevertheless the measurements showed that the three BBL domains unfold together and not within three repeatable unfolding patterns. Hence unfolding peaks were detected with the threefold value of ΔL of (60nm) than the corresponding value of one BBL domain (e.g see figure 5.8 panel B). This behavior is very different than the unfolding behavior of the titin I27, were each domain in the polyprotein construct unfolds individually.
As being observed during the measurements of $I27_3$-BBL-$I27_3$ construct, the change in the pulling velocity did not seem to change the general observations. Another interesting point is that the average value calculated for the unfolding force is very close to the corresponding value estimated for the single BBL domain. The following table 5.2 summarizes the results of the three BBL domains and for comparison also of the titin I27 domain at the three used velocities.

	BBL	
Force(pN)	67.5 ± 57.6	74.8 ± 52.2
ΔL(nm)	28.1 ± 10.4	22.2 ± 20.3
velocity (nm/s)	400	1700
	titin I27	
Force(pN)	200.0 ± 27.9	256.2 ± 31.6
ΔL(nm)	28.0 ± 0.6	28.4 ± 0.5
velocity (nm/s)	400	1700

Table 5.2: Overview of the measured force and difference in contour length ΔL of the three BBL domains and of I27

As these measurements showed a much higher complexity than the ones of the single BBL domain inside the $I27_3$-BBL-$I27_3$ construct no more experiments were conducted. At this point it was thought to use the pFS construct as it could clearly identify the BBL unfolding pattern inside the ubiquitin carrier.

5.1.3 Constant velocity measurements on the pFS-BBL construct

The pFS construct containing a single BBL domain was measured at a constant velocity of 400 and 1700nm/s using the MultiMode AFM. As this polyprotein construct contains the ubiquitin carrier, all traces were collected which showed its characteristic unfolding

length. In the case of the ubiquitin carrier with the BBL domain its complete unfolding length should be around 48nm, which is the length of the empty ubiquitin carrier and the BBL domain together (see also [174]).

Constant velocity measurements conducted at 400nm/s

Three force extension traces measured at a constant velocity of 400nm/s are shown in figure 5.10. All three depicted traces show the characteristic difference in contour length of the ubiquitin carrier containing the BBL domain as well as the complete unfolding of the pFS construct with the characteristic unfolding peaks of the five single ubiquitin domains. However instead of increasing the probability to clearly identify the BBL unfolding signal, within these measurements the unfolding signal of the BBL domain was found to be much more difficult to distinguish within the unfolding signal of the whole ubiquitin carrier, especially when using the WLC model. This issue resulted in a very high amount of traces, which showed no BBL unfolding pattern (see figure 5.10 A).

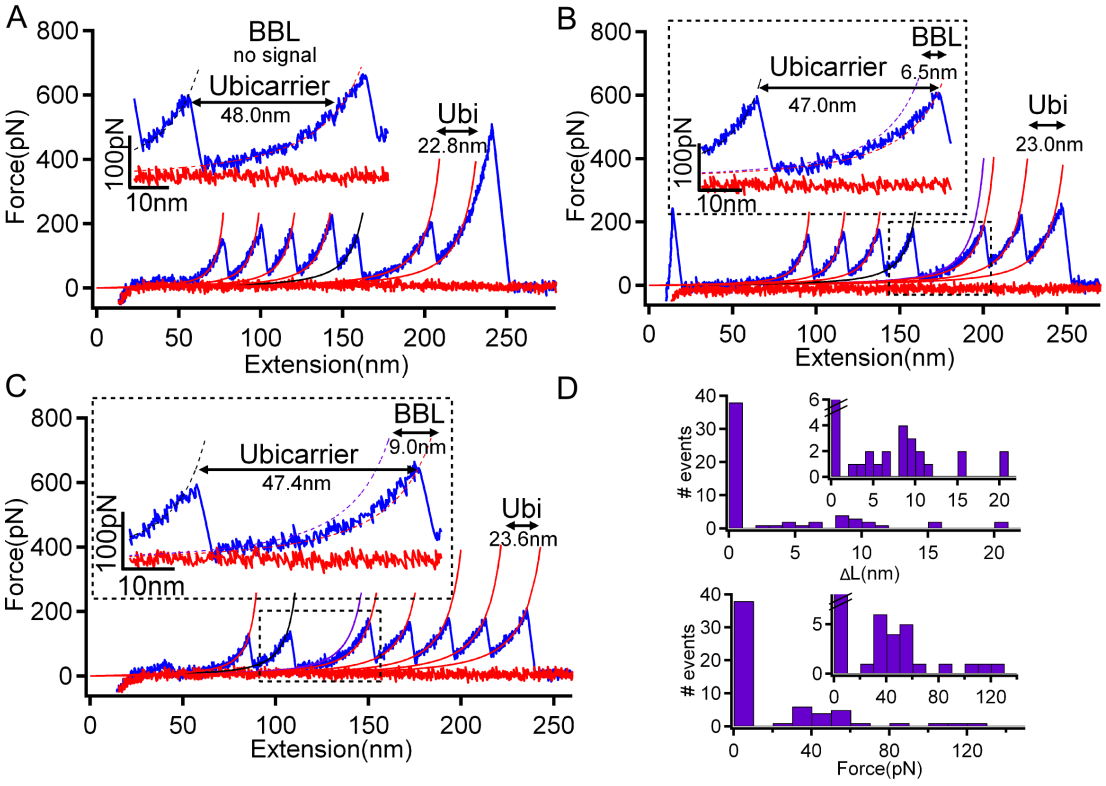

Figure 5.10: Force extension traces for the pFS ubiquitin construct containing the BBL domain at a constant velocity of 400nm/s (A) No BBL unfolding signal is detected, but a change in slope during the unfolding of the ubiquitin carrier (B) BBL unfolds through a very small unfolding peak (C) BBL unfolds though a very small unfolding peak/hump (D) Histograms of ΔL (1nm bin size) and unfolding force (10pN bin size) of the BBL domain inside the ubiquitin carrier

Here from a total amount of traces n=61, the big majority of around 62% (n=38) showed no or no distinguishable BBL unfolding signal. E.g. the traces in figure 5.10 A shows a change in slope at low force. The rest of traces around 39% (n=23) showed an distinguishable unfolding peak or unfolding hump for the BBL (see figure 5.10 B and C). Even in this cases the application of the WLC model fit was close to the resolution limit, which made a clear separation of the BBL unfolding signal from the unfolding signal of the ubiquitin carrier impossible in most of the traces. However the traces shown in panel B and C indicate a clear noisy region at the beginning of the unfolding signal of the unfolding of the ubiquitin carrier, which is supposed to correspond to the unfolding signal of the BBL domain. In all the insets shown in figure 5.10 the thickness of the WLC fit line was reduced in order not to overlay with the experimental trace.

Panel D in figure 5.10 gives an overview of the measured distributions of unfolding force and difference in contour length of the BBL. Averaging the distributions for the BBL results in a mean unfolding force of (56.5 ± 29.0)pN and in a mean difference in contour length ΔL of (9.5 ± 5.0)nm. This average was calculated without including the 38 traces in which no BBL signal was detected, as this would result in a very tilted distribution.

Figure 5.11 gives an overview of the measured distributions of unfolding force and difference in contour length additionally to the corresponding values of BBL also for the ubiquitin carrier and also of the single ubiquitin domain.

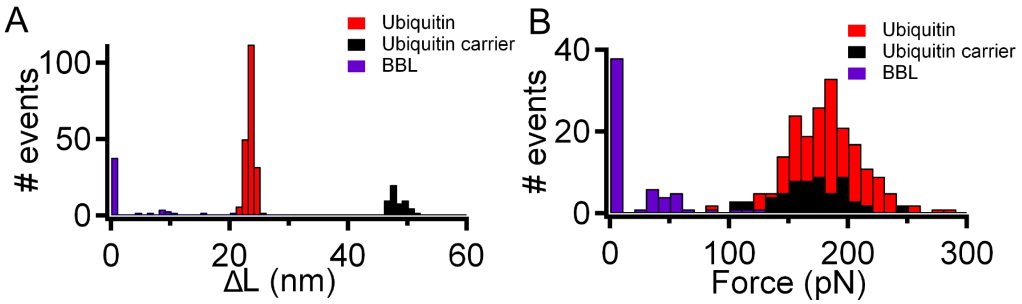

Figure 5.11: Histograms of the estimated difference in contour length (bin size 1nm) (A) and of the measured unfolding force (bin size 10pN) (B) of the BBL protein, the ubiquitin carrier and the single ubiquitin domains at a constant velocity of 400nm/s

Averaging the distributions the ubiquitin carrier results in an unfolding force of (169.2 ± 30.2)pN and in a difference in contour length including the length of BBL of (48.4 ± 2.3)nm. The corresponding results from the average of the single ubiquitin domains are (179.7 ± 32.2)pN and (23.3 ± 0.7)nm. The estimated values for the ubiquitin coincides very well with the corresponding values found in the reference [174].

Constant velocity measurements conducted at 1700nm/s

Measurements at a higher velocity of 1700nm/s unveil better distinguishable unfolding signal of the BBL domain inside the unfolding signal of the whole ubiquitin carrier. Figure 5.12 shows three typical force extension traces in which a unfolding signal of BBL could be identified.

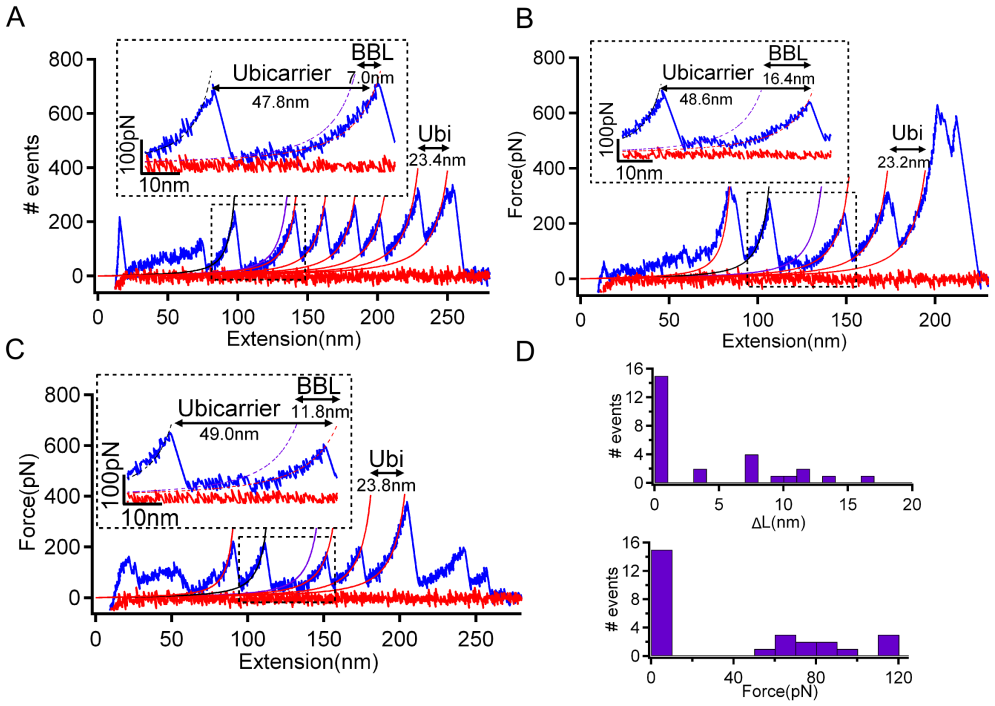

Figure 5.12: Force extension traces for the pFS ubiquitin construct containing the BBL domain at a constant velocity of 1700nm/s (A) BBL unfolding signal is detected (B) BBL unfolds through a small unfolding hump (C) BBL unfolds though a very clear unfolding peak (D) Histograms of ΔL (1nm bin size) and unfolding force (10pN bin size) of the BBL domain inside the ubiquitin carrier

The three traces in figure 5.12 A, B and C reveal this clearer BBL unfolding signal, although still 15 from a total amount of n=25 traces (60%) show no distinguishable BBL unfolding signal.

Panel D in figure 5.12 gives an overview of the measured distributions of unfolding force and difference in contour length of the BBL, although just n=25 traces have been measured. Averaging the distributions for the BBL results in a mean unfolding force of (83.9 ± 21.7)pN and in a mean difference in contour length of (8.9 ± 3.8)nm.

Figure 5.13 gives an overview of the measured distributions of unfolding force and difference in contour length for the BBL, the ubiquitin carrier and for the single ubiquitin domain. For the ubiquitin carrier an unfolding force of (221.0 ± 36.1)pN and in a difference in contour length ΔL including the length of BBL of (48.0 ± 1.2)nm has been estimated.

The corresponding results of the single ubiquitin domains are (226.2 ± 36.1)pN and (23.4 ± 0.7)nm.

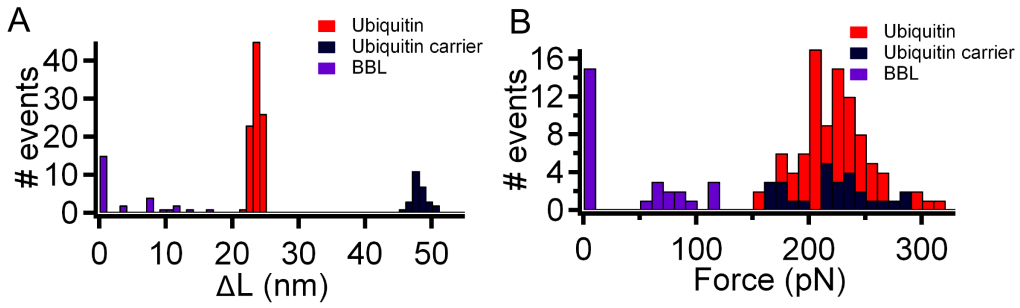

Figure 5.13: Histograms of the estimated difference in contour length (bin size 1nm) (A) and of the measured unfolding force (bin size 10pN) (B) of the BBL protein, the ubiquitin carrier and the single ubiquitin domains at a constant velocity of 1700nm/s

Summary and Discussion

The advantage of using the pFS ubiquitin vector is that the protein of investigation, in this case the BBL domain, always unfolds after the ubiquitin carrier unfolds. I.e the lower unfolding force signal from BBL does not appear at the beginning of the force extension trace (see e.g the corresponding force extension traces of the $I27_3$-BBL-$I27_3$ construct in figure 5.2). Therefore it cannot be mistaken by any interfering noise signal, which could be caused by any interaction of the cantilever with the surface. Hence the pSF construct should offer an advantage in measuring the BBL domain.

However the measurements showed that the unfolding signal of the BBL domains was hard to distinguish from the corresponding force extension traces at least when using the WLC model to fit the force extension trace. Therefore the majority of traces (60%) measured at both velocities did not reveal any BBL unfolding signal. It has been found that when increasing the speed to 1700nm/s the BBL signal appeared more distinctive, although the percentage of traces revealing no BBL signal was still the same. Table 5.3 gives an overview of the measured force of the BBL inside the pFS vector.

	BBL	
Force(pN)	56.5 ± 29.0	83.9 ± 21.7
ΔL(nm)	9.5 ± 5.0	8.9 ± 3.8
velocity (nm/s)	400	1700

Table 5.3: Overview of the measured force and difference in contour length ΔL of BBL inside the pFS vector

These values were estimated without counting the amount of traces in which no BBL signal was detected. Therefore a direct comparison with the corresponding values obtained for the $I27_3$-BBL-$I27_3$ construct (see table 5.1) cannot be made. In sum the application of the pFS system on the BBL domain did not show further details of its unfolding behavior. The reason for this high amount of traces showing no signal could be that the BBL domain inside the ubiquitin carrier is more constrained to be in the nearly unfolded state or the unfolding signal of the BBL cannot be distinguished from the unfolding signal of the ubiquitin carrier with the WLC-model as its unfolding length and unfolding force is one the resolution limit.
In order to increase the probability to detect a distinguishable unfolding signal of the BBL domain inside the pFS vector, the $\mathrm{pFS-BBL_2}$ construct was designed and tried as a last intention for using the pFS system for the BBL domain.

5.1.4 Constant velocity measurements on the $\mathrm{pFS-BBL_2}$ construct

With this construct just experiments at a constant velocity of 1700nm/s were conducted, because measurements at this velocity indicated a clearer signal of the inserted BBL domain in the measurements of the pFS-BBL construct. As for the measurements on the pFS-BBL construct before, all traces were collected which showed the characteristic unfolding length of the ubiquitin carrier. In the case of the $\mathrm{pFS-BBL_2}$ construct this length should be around 68nm, which is the length of the empty ubiquitin carrier and the two inserted BBL domains together.
Three typical force extension traces are depicted in figure 5.14. All three depicted traces show the characteristic unfolding length of the ubiquitin carrier containing the BBL dimer. In general the unfolding signal of the BBL dimer ($\mathrm{BBL_2}$) was more distinct and therefore better distinguishable than in the corresponding measurements of the pFS-BBL construct (see figure 5.12).
Although still n=18 from n=31 traces (60%) showed no BBL unfolding signal like shown in figure 5.14 A, the signals detected in panel B and C show a much clearer BBL unfolding signal, which can be fitted to the WLC-model. Interestingly, as it was also detected when measuring the $I27_2$-BBL-I27-BBL-I27-BBL-$I27_2$, the detected difference in contour length of the detected unfolding peak was larger than coming from one single BBL domain (see figure 5.14 B and C) and reached in some cases the length value of both BBL domains together.
Figure 5.14 D gives an overview of the measured distributions of unfolding force and

difference in contour length ΔL of the two BBL domains. The distributions for the BBL give a average value in unfolding force of (64.8 ± 23.0)pN and in ΔL of (18.9 ± 10.9)nm.

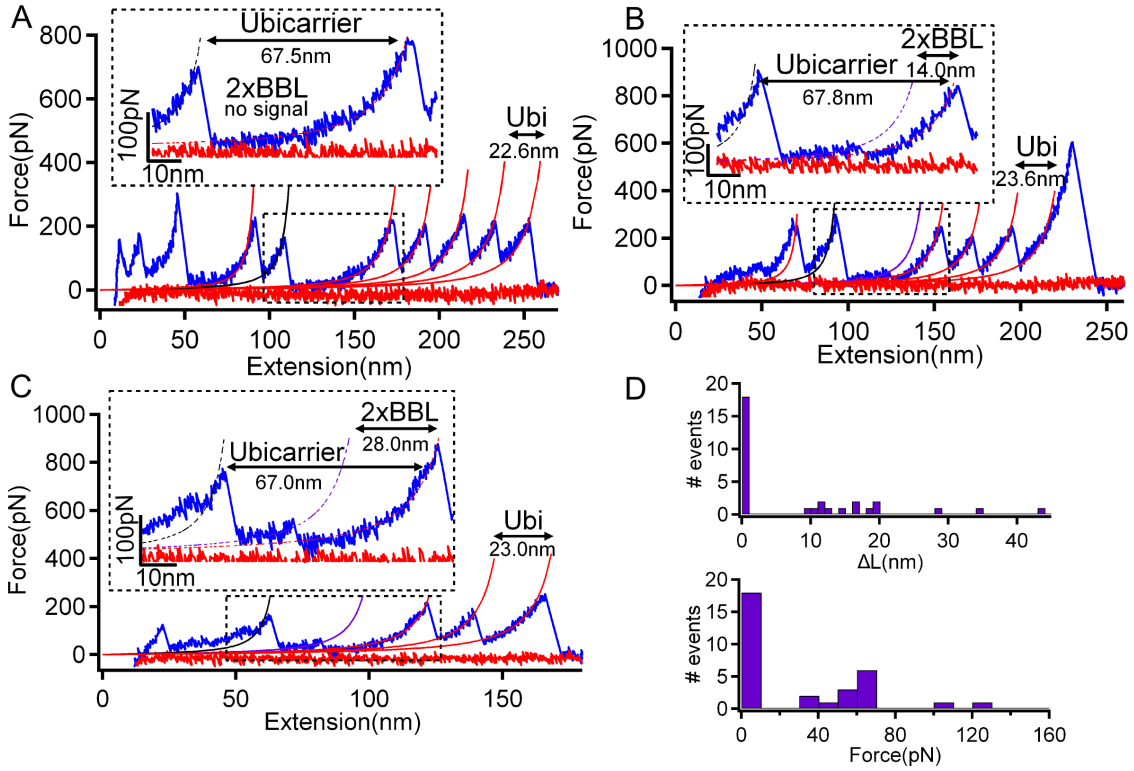

Figure 5.14: Force extension traces for the pFS ubiquitin construct containing the BBL dimer at a constant velocity of 1700nm/s (A) No BBL unfolding signal is detected, but a change in slope during the unfolding of the ubiquitin carrier (B) BBL unfolds through a unfolding peak (C) BBL unfolds through a very clear unfolding peak/hump (D) Histograms of ΔL (1nm bin size) and unfolding force (10pN bin size) of the BBL domain inside the ubiquitin carrier

The corresponding distributions for the ubiquitin carrier and for the single ubiquitin domain are shown in figure 5.15 together with the distributions measured for the two BBL domains from figure 5.14 D.

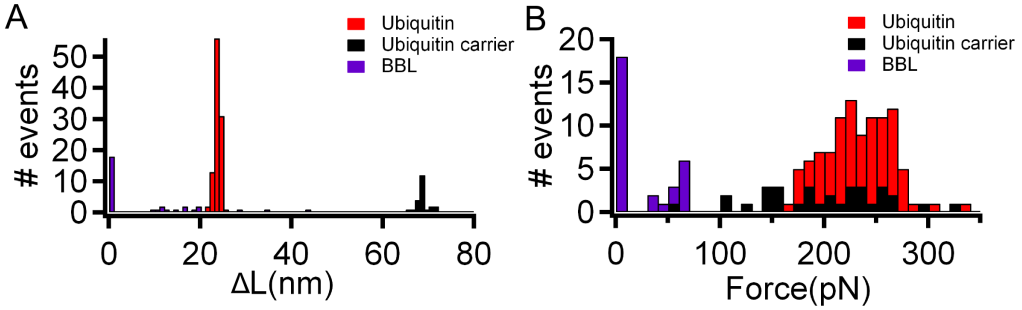

Figure 5.15: Histograms of the estimated difference in contour length (bin size 1nm) (A) and of the measured unfolding force (bin size 10pN) (B) of the two BBL domains, the ubiquitin carrier and the single ubiquitin domains at a constant velocity of 1700nm/s

Averaging the distributions the ubiquitin carrier results in an unfolding force of (198.8 ± 60.4)pN and in a difference in contour length including the length of the two BBL domains of (68.5 ± 1.3)nm. The corresponding results from the average of the single ubiquitin domains are (230.6 ± 33.1)pN and (23.6 ± 0.6)nm.

Summary and Discussion

Inserting the BBL dimer into the ubiquitin carrier affirmed a improved detectable unfolding signal of BBL domain. Furthermore it also confirmed the observation that the adding of single BBL domains results in a signal which contains both domains, resulting in a longer difference of contour length which coincided with the sum of the corresponding length of both individual BBL domains.

However the complexity of the measured unfolding patterns and the high amount of traces which revealed no detectable BBL signal impedes a further analysis of the traces. The last intention of measuring the BBL domain with the MultiMode AFM was with using the BBL polyprotein construct, which consisted of six BBL domains.

5.1.5 Constant velocity measurements on the BBL_6 construct

Measurements on the BBL polyprotein construct were performed at a constant velocity of 400nm/s. As no molecular fingerprint neither the one of the I27 titin domain nor of ubiquitin carrier is present in this vector, just the total unfolding length was be used to select the traces. The total unfolding length of the BBL_6 construct is around 120nm, which results as each BBL domain consists of 50 amino acids and that the length of one amino acid is approximately 0.4nm. Two force extension traces are shown in figure 5.16.

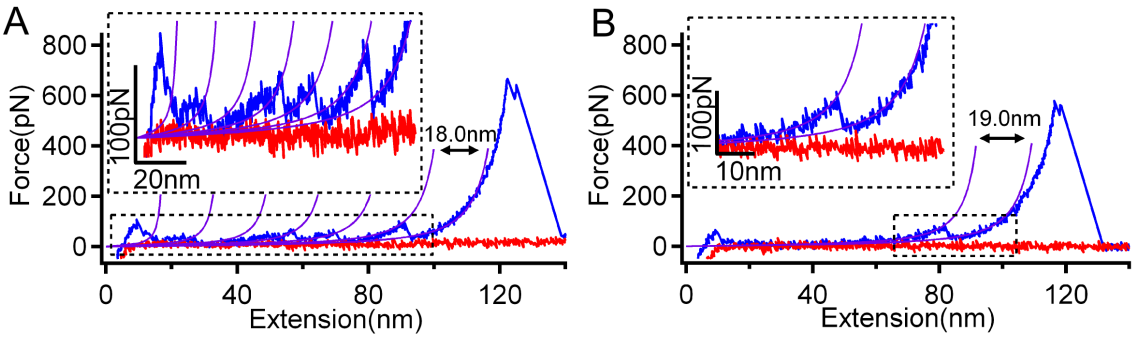

Figure 5.16: Force extension traces for the BBL_6 construct at a constant velocity of 400nm/s (A) Multiple unfolding signals from the six BBL are visible (B) One unfolding peak of the six BBL is visible

Both examples show a force extension trajectory, in which the cantilever gets deflected after around 120nm showing the huge disruption peak, which is correlated to the detach-

ment of the cantilever from the sample. Whereas figure 5.16 A indicates some unfolding signals before the detachment peak, figure 5.16 B shows just one unfolding peak before the disruption peak. In both traces WLC fittings indicate that the small peaks appear roughly within a difference in contour length of 18-19nm, which corresponds to the total length of the BBL domain used including its tails (see also figure 3.2).

Summary and Discussion

The BBL_6 construct was not showing a reliable unfolding pattern. The main difficulty was to achieve clean traces with no interaction signal at the beginning of the trace. As this could not achieved throughout the measurements it is suggested that the BBL_6 sample could agglomerate and also diluting the sample is not improving the aspect of the force extension traces.
Hence it is very difficult to trust the obtained force extension measurements with the BBL_6 construct and further attempts to measure this sample were omitted.

5.2 SMD simulations on the BBL domain

The two conducted SMD simulations at a constant velocity started from the same MD equilibrium simulation. Therefore the first section will provide the MD results. Then the constant velocity and additionally performed constant force SMD simulations will be shown. The 3Dim structure of BBL is shown in figure 5.17 depicting the two alpha helices α_1 and α_3, the small alpha helical turn α_3 and the N and C terminus to have a better orientation for the following sections.

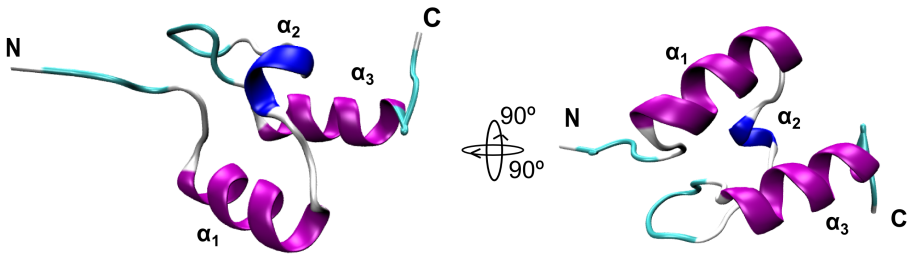

Figure 5.17: The 3Dim structure of the BBL is shown and the structural elements are indicated.

5.2.1 MD equilibrium simulation

Before the MD equilibrium simulation was started, the waterbox containing the BBL needed to be designed to adopt the experimentally used sequence.

First the pdb file 2cyu.pdb was downloaded and used for the first SMD simulations. This structure was then solvated into a waterbox with the dimensions (x y z) of (0 0 0) (23 5.5 5.5)nm. The final waterbox contained 66242 atoms. Afterwards the BBL structure was minimized using steepest decent for 0.1ns and equilibrated for 1ns.
However in order to have a identical and comparable structure to the experimental one the pdb file 2wxc.pdb was downloaded as it contains the tails of the protein. However this structure needed to be manipulated. E.g the containing tryptophan (W 142) needed to be exchanged to a histidine (H 142), which is present in the experimental used structure. Furthermore the downloaded structure the C-terminal was cut by the first amino acid Q. Then at the C-terminal two aminoacids KK and at the N-terminal four amino acids PAKK were added. Finally a waterbox was constructed with the dimension (x y z) of (0 0 0)(23 5.5 5.5)nm. This final waterbox contained 66264 atoms. Afterward this BBL structure inside the waterbox was minimized using steepest decent for 1ps and equilibrated for 0.3ns. After the simulation was run for both structures, the RMSD value was calculated for each timestep of the equilibrium simulation, which compares the position of the backbone atoms of the original and the simulated structure and which is shown in figure 5.18.

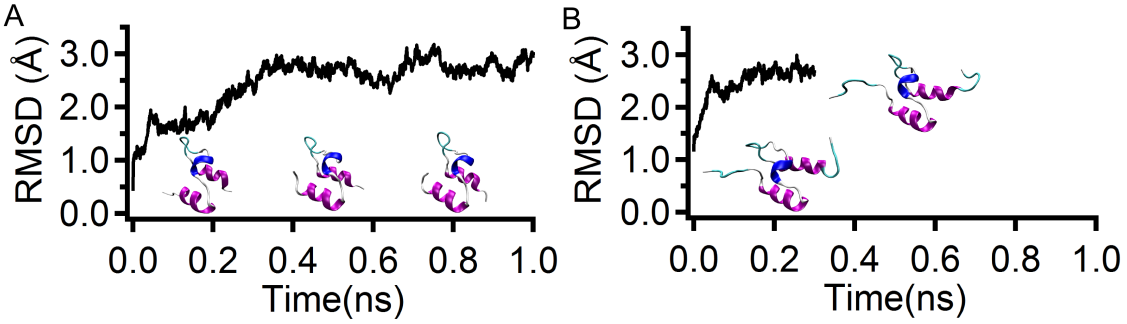

Figure 5.18: Change in the RMSD value during the MD simulation. Corresponding snapshots of the short BBL structure (2cyu.pdb)(A) and of the designed BBL structure (B) are shown.

As it can be seen in the panel B, after 0.3ns the structure already found its stable position.

5.2.2 Constant velocity SMD simulations

Two SMD force/extension curves at a constant velocity of 1m/s is shown in figure 5.19. Panel A shows the constant velocity SMD simulation trace on the short BBL domain structure (2cyu.pdb) and panel B shows the corresponding SMD simulation force extension trace of the designed longer BBL structure. Although one simulation is not enough to do statistical predictions of the BBL computational unfolding behavior, it can still reveal a basic idea. The force extension curves panel A and B plots the force applied on the C_α-

atom of the C-terminal residue and its change in x-direction during the application of a movement with a constant velocity in x-direction.

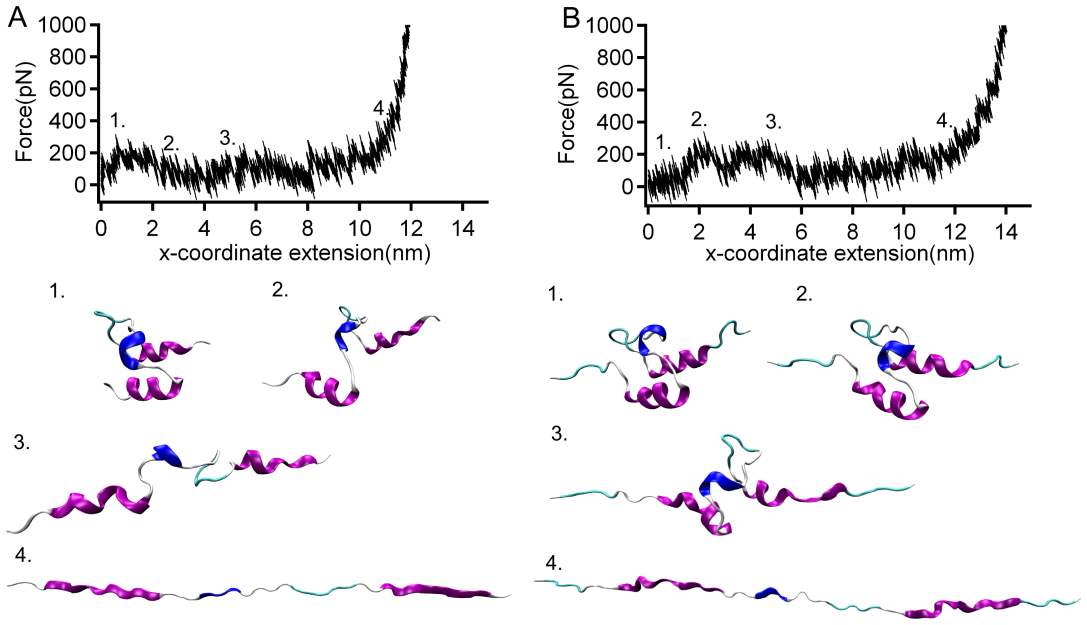

Figure 5.19: A) SMD simulation at a constant velocity of 1m/s of the short BBL structure and its corresponding snapshots (B) SMD simulation at a constant velocity of 1m/s of the longer and experimentally used BBL structure and its corresponding snapshots

The lower part of each panel depicts the snapshots from the corresponding SMD trajectory at the given point numbered in panel A and B from 1 to 4. The x-coordinate extension of the pulling C_α atom was set to zero in both force/extension curves, when the SMD simulation starts in order to have a better comparison to the experimentally estimated BBL unfolding length.

The unfolding force of both BBL structures is low and reaches around 200-250pN at the beginning of the trace. For both BBL structures no certain force bearing peak is visible, moreover the unfolding force seems to be regularly distributed throughout the whole trajectory. Interestingly the simulation done with the short BBL version (panel A) seems to unfold under a slightly less mechanical force than in the simulation done with the designed and tail containing BBL structure (panel B).

The snapshots of the corresponding BBL unfolding trajectory reveal a shearing of the two parallel helices α_1 and α_3 from point 1 (1ns, 0.6nm) to point 2 (2ns, 1.7nm) for the short BBL structure and from point 1 (1ns, 0.8nm) to point 3 (5ns, 4.7nm) for the longer BBL structure. At point 4 (12ns, 11nm and 12.7nm respectively) both BBL structures are fully stretched. An interesting difference between both structures is that for the longer structure the tail at the C-terminus seems to contain a force bearing connection to the

unstructured loop between alpha helix α_2 and alpha helix $\alpha 3$. Therefore between point 1 and point 2 (panel B) this connection breaks, before between point 2 and 3 the two parallel alpha helices α_1 and α_3 start to shear.

5.2.3 Constant force SMD simulations

Figure 5.20 shows all the BBL separation length vs time trajectories from the conducted constant force SMD simulations.

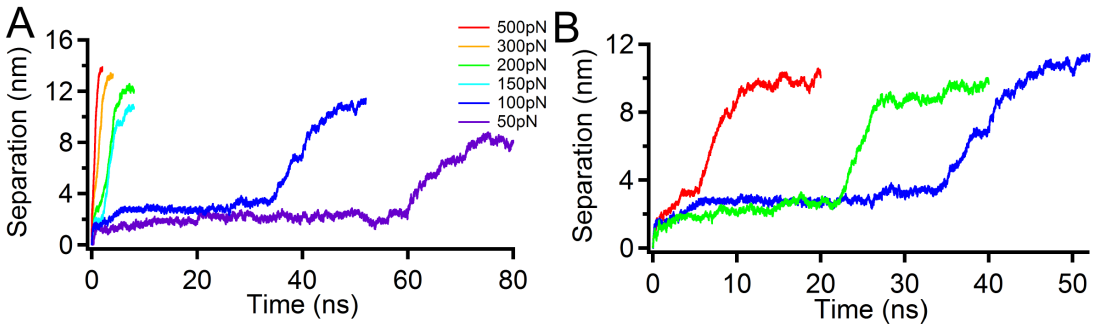

Figure 5.20: (A) Constant force SMD simualtions at 50, 100, 150, 200, 300 and 500pN (B) Three constant force SMD simulations at 100pN

In panel A five trajectories from SMS simulations at different constant forces 50, 100, 150, 200, 300 and 500pN are compared. Here it has been found that 100pN is an ideal force to conduct SMD simulations on BBL, as it unfolds in a reasonable calculation time, which can be performed by the used computational resources (see blue trace). Panel B shows the variety of three separation length vs time traces, all conducted at a constant force of 100pN. All BBL unfolding trajectories show a very similar pathway. In general just the two alpha helices α_1 and α_3 shear, become separated and then totally stretched.

Figure 5.21 illustrates the details of the blue trajectory at a constant force of 100pN as one example from figure 5.20 (A) and (B). Panel A shows the single separation vs time trajectory, which indicates a two step unfolding behavior, the first until 6ns and the second around 33ns. The overall unfolding time is 52ns and the corresponding unfolding length is around 11.4nm. Panel B shows the corresponding secondary structure content plotted using the timeline plugin of the VMD program. Together with the snapshots shown in panel C, the whole unfolding trajectory can be followed. The unfolding pathway of this trajectory is very similar to the one observed during the constant velocity simulation (see figure 5.19 panel B). The unfolding starts by stretching the tail at the C-terminus between 1 and 6ns. Afterwards the two parallel alpha helices α_1 and α_3 start to shear and keep in this position until 33ns.

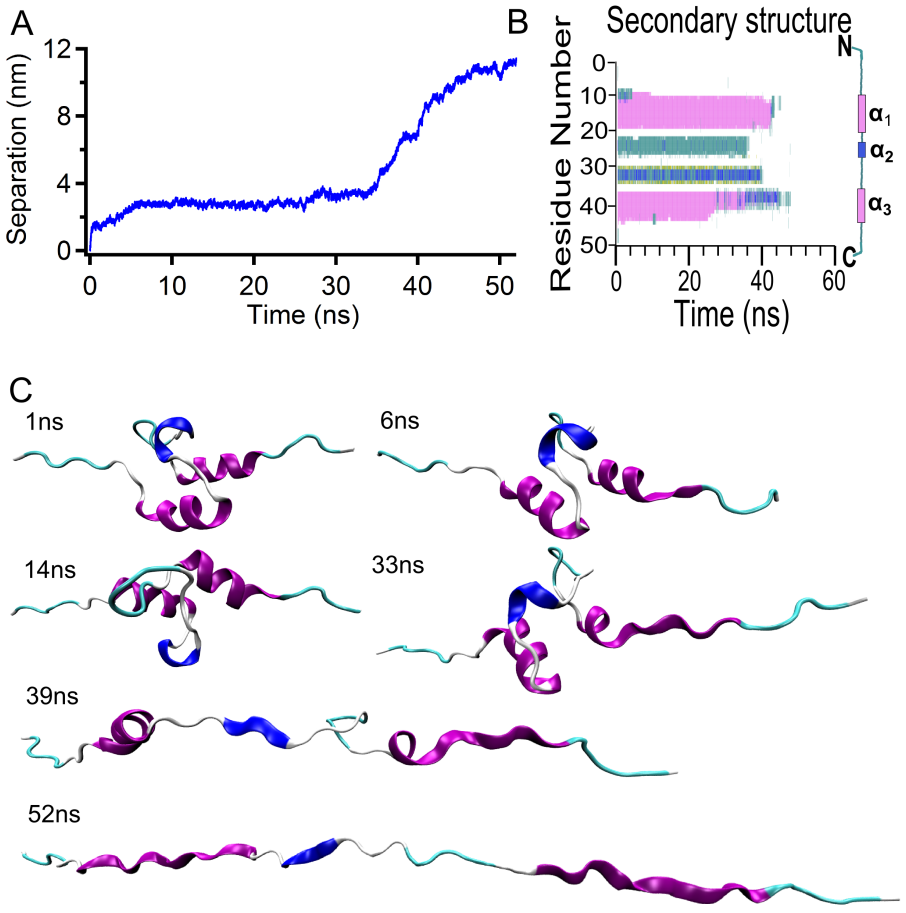

Figure 5.21: A) Constant force SMD simualtions at 100pN on the BBL structre B) Secondary structure content during the simulation C) Snapshots from the trajectory of the BBL, taken at the indicated time.

Interestingly, while being in this position the whole protein can turn around the pulling axis as it can be seen in the snapshots of the trajectory going from 6ns to 14ns and 33ns (panel C). At the point of 33ns then the two helices start to separate and disentangle one after another until the BBL protein is fully stretched.

Summary and Discussion

The conducted SMD simulations describe the unfolding pathway of the BBL domain at atomic resolution. Therein one simple unfolding pathways has been identified. All detected unfolding pathways at a constant force start with the disconnection of the tail of the C-terminus from the hydrophobic unstructured loop between the two alpha helices α_2 and α_3. Then the two parallel alpha helices α_1 and α_3 start to shear until they become fully stretched one after another.

The performed SMD simulations reveal also a total difference in contour length of the BBL structure to be around 13-15nm.

5.3 Conclusions

Investigating the mechanical response of the fast folding BBL domain in constant velocity mode revealed a highly complex and heterogeneous system, which additionally is at the detection or resolution limit of the MultiMode AFM. Hence forces below 15-20pN cannot be detected. All the attempts to improve the detection of the BBL unfolding signal in designing different polyprotein constructs showed in some cases some advances (e.g measurements performed on the $I27_2$-BBL-I27-BBL-I27-BBL-$I27_2$ and the $pFS - BBL_2$ constructs) but on the same hand also an increase in complexity. The usage of the pFS construct was neither improving the measurements in the case of the BBL domain, because less signal was visible than compared to the $I27_3$-BBL-$I27_2$ construct. Furthermore the BBL polyprotein construct BBL_6 was failing in producing any improvement, instead the corresponding results were impossible to interpret.

Thus after unveiling this complex behavior of the BBL domain during the constant velocity measurements, a new strategy was chosen to investigate the mechanical unfolding behavior of fast folding proteins.

First a clear 2-state, mechanical force resistant and moderate fast folding protein was chosen to be investigated in order to ensure that the performed AFM experimental and used molecular biology polyprotein sample building strategies were correct. Here the cold shock protein from *Thermotoga Maritima* (Csp) has been found in the literature, which fulfills all the mentioned conditions (as explained in chapter 3.3). The advantage of choosing the Csp as a comparison was that it had been already measured under a mechanical force using the constant velocity mode of the AFM (see [128]), wherein it showed to unfold via a clear single peak at a force of around 78pN at a velocity of 400nm/s. Furthermore the Csp has been described to have a folding rate $k_f \approx 565s^{-1}$, which is in the moderate fast folding regime [186].

Second a fast folding and over a marginal energetic barrier folding protein was chosen to be investigated in order to minimize the already discovered complexity in the BBL. Therefore the gpW protein was selected for the following reasons (see also chapter 3.4). It folds in a very close time regime to the BBL domain, which lies in the lower 100-50μs range (40.000-20.000s^{-1} [78]). Beside two α helices it contains also two β-strands, which is thought to increase the mechanical response compared to the one from the BBL. Furthermore the gpW protein contains also more residues than BBL, which would improve the detection of the unfolding signal.

Finally, additionally to the MultiMode AFM, the force-clamp AFM was used in order to

improve the resolution in time, force and length. The force-clamp AFM also enabled to measure the new chosen samples under a controlled force ramp and constant force. These measurements were performed in the Nanobiomechanics group of Dr. Raul Perez Jimenez at the nanoGUNE institute in San Sebastian (Spain).

Chapter 6

Mechanical unfolding of the cold shock protein (Csp)

This chapter presents the overall results describing the mechanical behavior of the Csp, which will be divided into the following sections: The first section covers all the conducted SMFS experiments on the Csp, from constant velocity, force ramp to constant force AFM measurements. The kinetic stochastic simulations of the Csp will be described in the end of the performed SMFS experiments section. The second section describes the results of the conducted SMD simulations on the Csp.

6.1 SMFS measurements on the Csp

For all Csp measurements just traces were collected, in which at least four titin I27 unfolding events could be detected, which was already discussed in detail in the BBL results chapter. Furthermore as explained before (see figure 5.1) traces were discarded, which showed multiple polyprotein adsorption or too much noise at the beginning of the trace. In the case of the Csp, the range in unfolding length before the first titin I27 domain unfolds changes a bit as the Csp consist of more amino acids than the BBL. The expected unfolding length of the Csp is around 26.4nm when multiplying its number of amino acids 66 by the approximated length of a single amino acid, which is 0.4nm. Using again the estimated length of the six folded titin I27 domains (26.4nm), the irst I27 domain should unfold after a length between 26.4nm and 52.8nm. Examples of discarded traces will be given in each section for the force ramp and the constant force mode respectively, as this kind of measurements has not been discussed before in order to explain which conditions have been applied to select good and usable traces followed throughout this work.

6.1.1 Constant velocity measurements

Also for the Csp, in all constant velocity measurements at the beginning of the trace the cantilever was pushed first with $|\tilde{F}| = 1$nN against the surface, before the cantilever is retracted. With that amount of force protein adsorption to the cantilever was ensured. As it has been shown in the chapter of the BBL, each force/extension trajectory the red trace represents the approach and the blue trace the retraction of the cantilever from the surface. The insets in each trajectory figure show the beginning of the trace at a higher magnification to highlight the unfolding behavior of Csp, because Csp has a lower unfolding force (see [128]) than the titin I27 domain.

Constant velocity measurements at 400nm/s

Two typical force extension traces measured with the MultiMode AFM at a constant velocity of 400nm/s are illustrated in figure 6.1. Both depicted traces show the complete unfolding of the polyprotein construct, because the Csp and the six titin I27 unfolding peaks are visible.

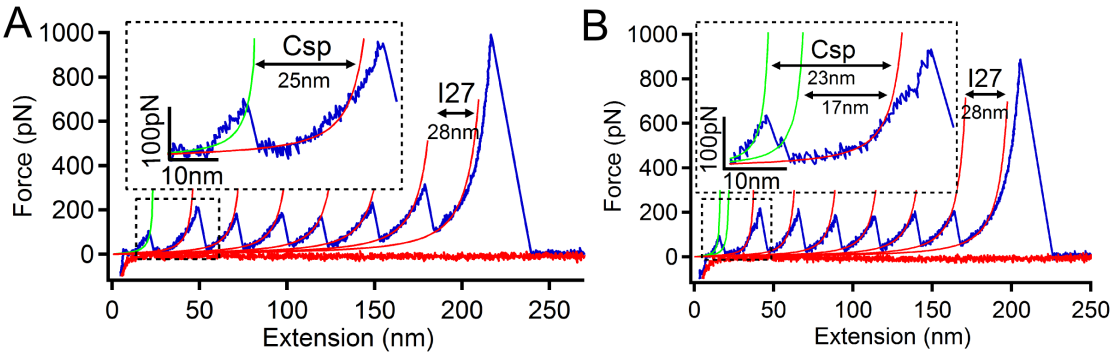

Figure 6.1: Force extension traces for Csp obtained a constant velocity of 400nm/s (A) Csp unfolds through a single peak (B) Csp unfolds through a single peak but with a second hump indicated with WLC model fit.

The majority of all selected traces (n=52) showed a single unfolding peak of the Csp (see figure 6.1 A). Therein a difference in contour length of around 25nm has been measured. However in up to 10% of the traces (n=5) the detected Csp unfolding behavior differed from a single unfolding peak (see panel B). Here for example the Csp unfolding peak shows a second hump with an intermediate a difference in contour length of around 17nm. However in general, the unfolding signal from the Csp measured at 400nm/s was very clear distinguishable and good traces could be easily detected compared to the corresponding measurements at the single BBL domain (see figure 5.2).

Figure 6.2 clarifies the measured distributions of unfolding force and difference in contour length of the Csp and for of the titin I27 domain, after fitting each unfolding peak to the WLC model.

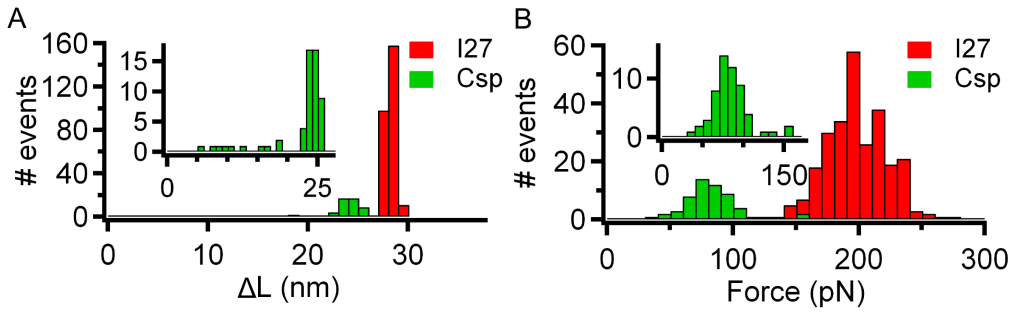

Figure 6.2: Histograms of the distribution of the estimated difference in contour length (bin size 1nm)(A) and the measured unfolding force (bin size 10pN) (B) of the Csp and the titin I27 domain at a constant velocity of 400nm/s

Averaging the distributions for the Csp results in a mean unfolding force of (82.4 ± 22.6)pN and in a difference in contour length of (24.1 ± 0.9)nm. The corresponding results from the average of the titin I27 domain are (197.6 ± 25.6)pN for the unfolding force and (28.1 ± 0.5)nm for ΔL. All the estimated values coincide perfectly with the corresponding values found in the literature [128]. However as some traces revealed a possible multiple peak unfolding mechanism of the Csp, which has not been reported before, constant velocity measurements were conducted at decreased speed of 10nm/s.

Constant velocity measurements at a velocity of 10nm/s

Figure 6.3 shows three typical force extension traces at a constant velocity of 10nm/s, which were used for the data analysis. All three traces show the Csp and the six titin I27 unfolding peaks. Hence the whole polyprotein construct was completely unfolded. The explanation of the higher noise in force of the red approach trace (± 9pN std compared to ± 5pN std of the blue retraction trace) in figure 6.3 is that during these experiments the approach velocity of the cantilever was set to 400nm/s and just the retract velocity to 10nm/s in order to save measuring time. It has to keep in mind that for a retraction cycle of a total length of 300nm at a velocity of 10nm/s, 30s are needed to complete the blue retraction trace. In contrast at a velocity of 400nm/s just 1-2 seconds are needed to measure one force extension trajectory. That makes these measurements at 10nm/s much more sophisticated compared to measurements at 400nm/s, as the AFM setup needs to be stable during 32s.

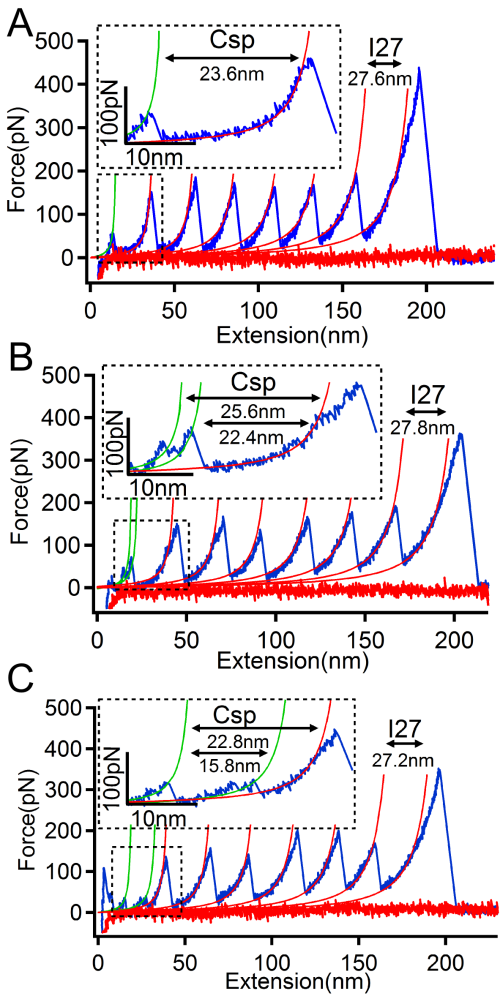

Figure 6.3: Force extension traces for Csp measured at a constant velocity of 10nm/s (A) Csp unfolds via one force peak (B) Csp unfolds via double force peak (C) Csp unfolds through 2 distinguishable force peaks

The force extension traces depicted in figure 6.3 show that the Csp unfolds via a single unfolding peak (panel A), however in the traces depicted in panel B and C the Csp unfolds via at least 2 unfolding peaks. This heterogeneity was not reported before [128] and was observed here for the first time. In comparison to the measurements done at a velocity of 400nm/s (see fig 6.1), at 10nm/s around 38% (n=12) of all traces (n=32) reveal a Csp unfolding behavior different from a single unfolding peak.

Figure 6.4 depicts the corresponding distributions of unfolding force and difference in contour length of the Csp and of the titin I27 domain. The sum of the intermediate peak lengths is indicated in the inset of panel A as the white distribution, which is around 25nm. This sum in length was then used to estimate the average value of the difference in contour length of the Csp. Averaging the distributions for the Csp results in a mean difference in contour length of (23.6 ± 0.9)nm and in a mean unfolding force of (57.0 ± 22.4)pN. The corresponding results from the average of the titin I27 domain are (27.4 ± 0.4)nm and (158.5 ± 22.5)pN. The value of the I27 domain coincides also well with the values found in the literature at this low speed of 10nm/s [167].

In general the experiments performed at 10nm/s revealed the capacity of the MultiMode AFM to conduct highly precise and stable measurements.

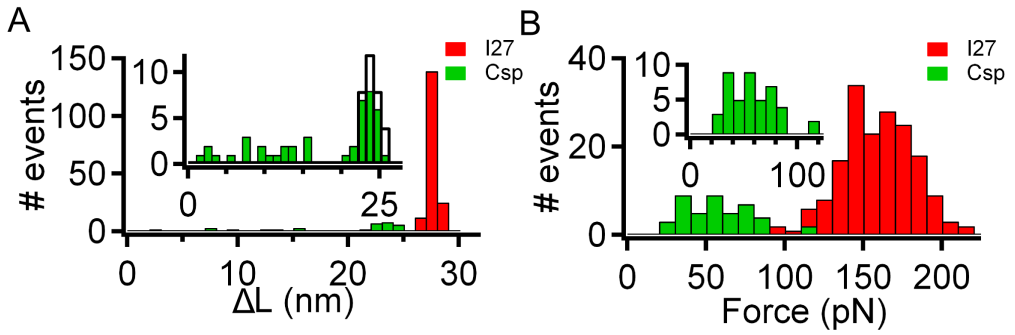

Figure 6.4: Histograms of the distributions of the estimated difference in contour length (bin size 1nm)(A) and the measured unfolding force (bin size 10pN) (B) of the Csp and the titin I27 domain at a constant velocity of 10nm/s

Summary and Discussion

The constant velocity data obtained at 400nm/s confirm the measurements published in the Lorna Dougan *et al* article [128]. Furthermore the force and the difference in contour length of the I27 titin domain was found to be in the same range, which refers to a well calibrated AFM MultiMode instrument, which also applies to the performed measurements on the BBL domain. Hence, this confirmation enabled to continue with the SMFS experiments. However, the new measurements at 10nm/s indicate a higher complexity of the mechanical unfolding behavior of the Csp. Here in 12 of 32 recorded traces the Csp unfolds at least via 2 peaks or two unfolding events. This indicates that maybe in a lower force range the Csp unfolds via intermediate states. Table 6.1 summarizes the results at both velocities.

	Csp		titin I27	
Force(pN)	82.4 ± 22.6	57.0 ± 22.4	197.6 ± 25.6	158.5 ± 22.5
ΔL(nm)	24.1 ± 0.9	23.6 ± 0.9	28.1 ± 0.5	27.4 ± 0.4
velocity (nm/s)	400	10	400	10
Nr. of traces (n)	52	32	52	32
% of intermediates	10	38	0	0

Table 6.1: Overview of the measured force and difference in contour length of Csp and of I27

As expected the mean of the unfolding force of Csp decreases when decreasing the velocity from 400 to 10nm/s, which applies also for the I27 domain and is again related to the effect of force and especially its amplitude change with time on the folding/unfolding rate on the protein (see Bell equation 2.3) as it has been also explained in the corresponding

measurements of the BBL domain. The table also displays the percentage of traces of the Csp, which were found to show an intermediate unfolding signal. As it has been mentioned before, this fact maybe indicates that with dropping the velocity, the probability of Csp to unfold via an intermediate is increased.
However as the constant velocity measurements are always coupled with noise signals at the beginning of the trace and lack of control/resolution (see also figure **??**), this possible unfolding behavior of the Csp through intermediate states is not clear distinguishable enough even at a small velocity like 10nm/s. Therefore the following force clamp measurements were conducted to provide more insight into the mechanical unfolding behavior of the Csp.

6.1.2 Force ramp measurements

In all force ramp measurements at the beginning of the trace the cantilever was pushed first with $|\tilde{F}| = 1$nN against the surface. However then in contrast to the before explained constant velocity measurements, this force value was then decreased within 0.1 s to $|\tilde{F}| =$ 20pN, from which the force was then linearly increased with the corresponding time rate up to a force value of $|\tilde{F}| = 300$pN in order to detect the titin I27 unfolding steps in the end of the trace. Here the linear increase corresponds to a force direction away from the surface (pulling).
This enables a cleaner transition from pushing to pulling. In each individual trajectory shown, the red trace represents the length vs time and the black the corresponding force vs time trace. Additionally, the estimated step length of the Csp and of the I27 titin domain are indicated in each trace. The insets show the corresponding unfolding step of the Csp at a higher magnification to focus on its unfolding behavior in particular. The spikes in the force vs time trace indicate the unfolding of a protein domain, where the feedbackloop needed to readjust the force in order to keep the set linearly increasing force constant. The amplitude of these force spikes depends strongly on each single cantilever chosen and therefore is sometimes higher or lower. The unfolding force and the corresponding length at each step was directly measured of the Csp and the titin I27 domain and in contrast to the constant velocity measurements not derived from any applied model.
As it was shown in the case of constant velocity measurements, figure 6.5 indicates two traces as an example of traces which needed to be discarded for the following reasons:
The trace in figure 6.5 (A) indicates the Csp unfolding signal (step) at low force after around 6s, however the unfolding of the I27 titin domains shows several irregularities.

Therefore similar traces were discarded, although the Csp signal could be detected.
In figure 6.5 (B) a trace is shown with the typical I27 titin unfolding pattern in the end, however there are some unfolding patterns at the beginning of the trace before the Csp unfolding pattern at around 2 and 4s. This would be another reason to discard the trace from the analysis.

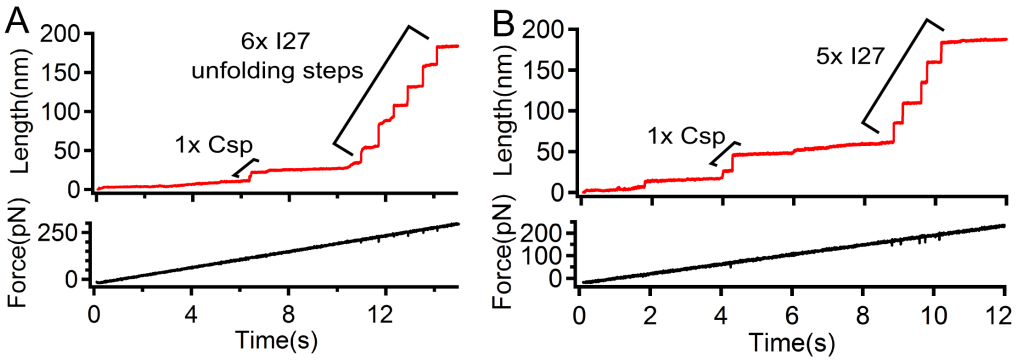

Figure 6.5: Two length and force vs time traces for Csp at a force ramp of 20pN/s as an example of traces which cannot be used for the data analysis (A)-(B). The unfolding steps of the Csp and I27 sare indicated

Force ramp measurements at a force rate of 20pN/s

Two typical selected force ramp traces at a constant force rate of 20pN/s are shown in figure 6.6. Both depicted traces show the complete unfolding of the polyprotein with the Csp and six titin I27 unfolding steps.

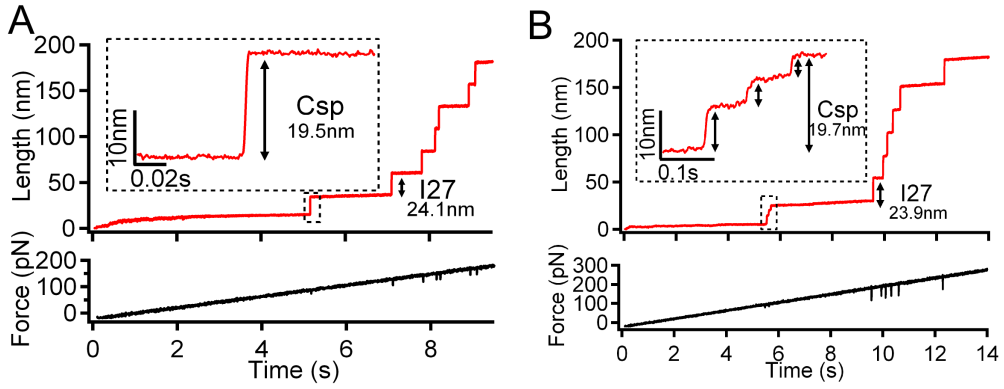

Figure 6.6: Force ramp traces for Csp at obtained at a constant force rate of 20pN/s (A) The Csp unfolds via one unfolding step (B) The Csp unfolds through three unfolding steps

As it can be seen the Csp unfolds via one steps (panel A), but additionally traces are detected in which the Csp unfolds via several clearly distinguishable steps. Panel B shows how the Csp unfolds via three steps. Around 45% of all traces (n=60) reveal a single step

unfolding behavior of the Csp, whereas in 55% Csp unfolds via several steps, which are indicated with small arrows in panel B.

When a trace was detected in which the Csp was unfolding via several steps, each step length was measured individually with its corresponding unfolding force and the sum of each individual step lengths was calculated. Figure 6.7 gives an overview of the measured distributions of unfolding force and step length of the Csp and for comparison also of the titin I27 domain.

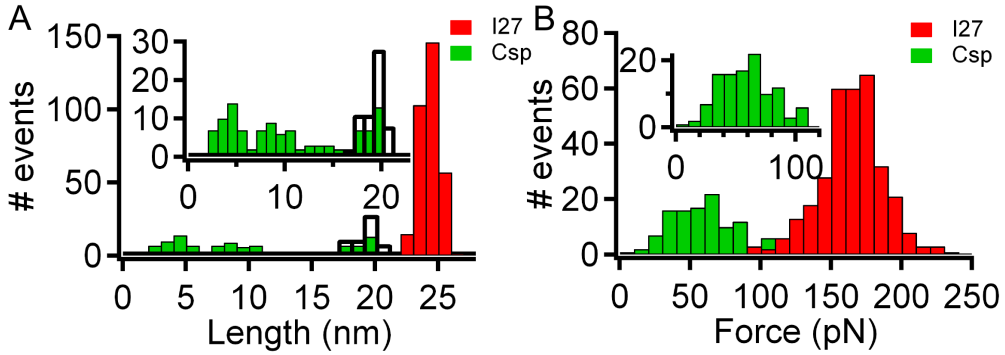

Figure 6.7: Histograms of the distributions of the estimated step length (bin size 1nm)(A) and the measured unfolding force (bin size 10pN) (B) of the Csp and the titin I27 domain at a constant force ramp of 20pN/s

The sum of the individual step lengths in the traces, in which Csp unfolds via several steps, is indicated in the inset of panel A as a the white distribution. This sum was then used to estimate the average value of the total step length of the Csp. For calculating the average of the unfolding force of Csp, all detected forces at each individual step length were used. Averaging the distributions for the Csp results in a mean step length of (19.0 ± 1)nm and in a mean unfolding force of (58.7 ± 22.2)pN. The corresponding results from the average of the titin I27 domain are (24.3 ± 0.7)nm and (164.5 ± 22.8)pN.

Force ramp measurements at a force rate of 80pN/s

Figure 6.8 shows two usually occurring force ramp traces at a constant force rate of 80pN/s. It has been observed that also at 80pN/s, the Csp can unfold via one unfolding step (see panel A) or via several steps, here like in panel B via four clearly distinguishable steps. Around 45% of all traces (n=22) reveal a single step unfolding behavior of the Csp, whereas in 55% Csp unfolds via several unfolding steps. Figure 6.9 shows the corresponding measured distributions of unfolding force and step length of the Csp and of the titin I27 domain. As before the white distribution in panel A shows the summed length in each trace of the Csp. Averaging the distributions for the Csp results in a mean step length

of (19.2 ± 1.0)nm and in a mean unfolding force of (62.4 ± 24.6)pN. The corresponding results from the average of the titin I27 domain are (24.5 ± 0.6)nm and (171.4 ± 29.1)pN.

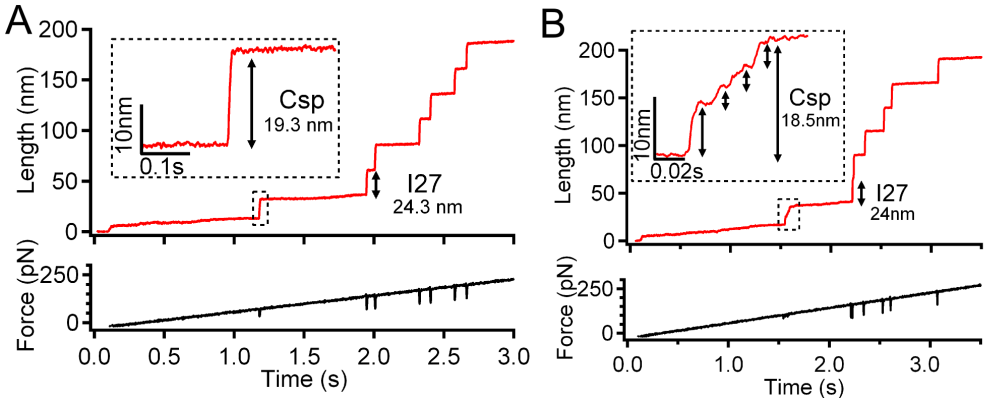

Figure 6.8: Force ramp traces for Csp measured at a constant force rate of 80pN/s (A) The Csp unfolds via one step. (B) The Csp unfolds via four steps

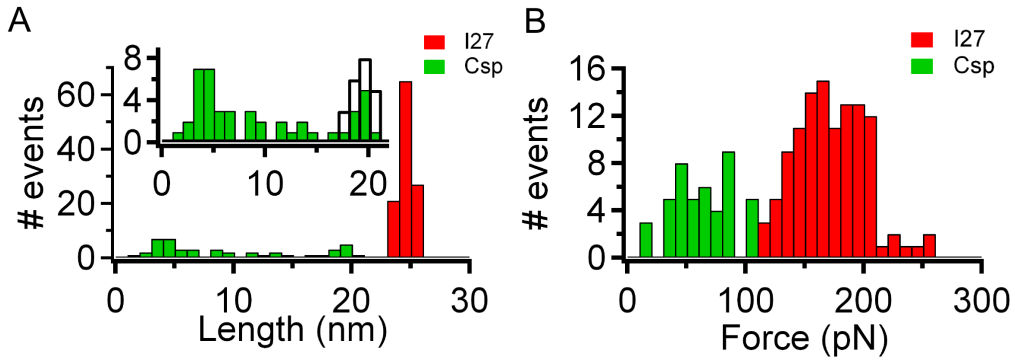

Figure 6.9: Histograms of the distributions of the estimated step length (bin size 1nm) (A) and the measured unfolding force (bin size 10pN) (B) of the Csp and the titin I27 domain at a constant force ramp of 80pN/s

Force ramp measurements at a force rate of 200pN/s

At a constant force rate of 200pN/s no noticeable changes are observed in the already observed unfolding behavior of the Csp at lower force rates (see figure 6.10). At 200pN/s again Csp can unfold via one unfolding steps (see panel A) or via several steps, here in panel B via five unfolding steps. Around 51% of all traces (n=63) reveal a single step unfolding behavior of the Csp, whereas in 49% Csp unfolds via several steps.

Figure 6.11 depicts the measured distributions of unfolding force and step length of the Csp and of the titin I27 domain. Averaging the distributions for the Csp results in a mean step length of (18.8 ± 1.1)nm and in a mean unfolding force of (67.8 ± 27.3)pN. The corresponding results from the average of the titin I27 domain are (24.5 ± 0.7)nm and

(200.6 ± 28.0)pN.

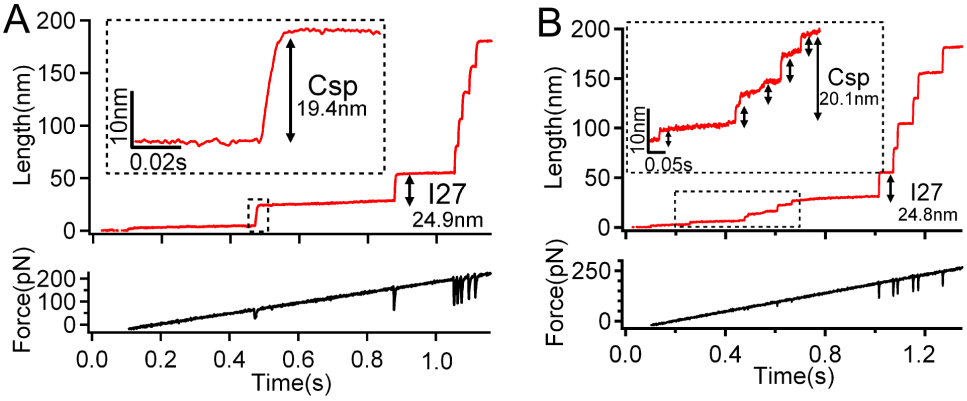

Figure 6.10: Force ramp traces for Csp obtained at a constant force rate of 200pN/s (A) The Csp unfolds via one single step (B) The Csp unfolds via five unfolding steps

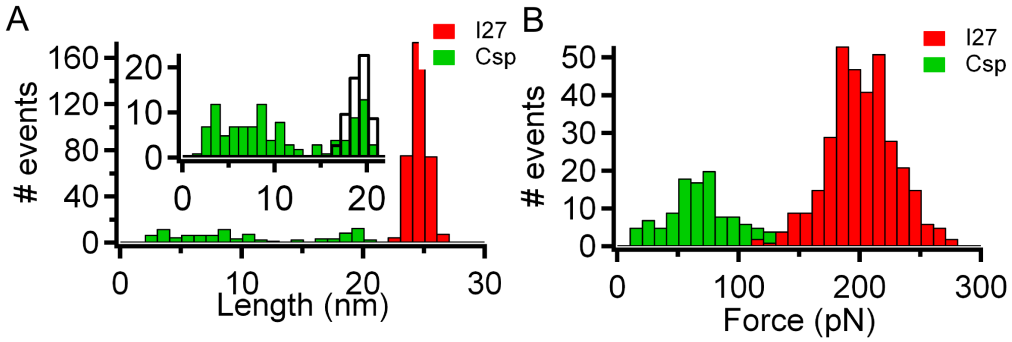

Figure 6.11: Histograms of the distributions of the estimated step length (bin size 1nm) (A) and the measured unfolding force (bin size 10pN)(B) of the Csp and the titin I27 domain at a constant force ramp of 200pN/s

Force ramp measurements at a force rate of 800pN/s

Figure 6.12 shows that also at a force ramp rate of 800pN/s, the Csp can unfold via one (panel A) or several steps (panel (B). I.e. even at this higher speed the before described Csp unfolding behavior does not change. Around 55% of all traces (n=56) reveal a single step unfolding behavior of the Csp, whereas in 45% Csp unfolds via several steps. The measured distributions of unfolding force and step length of the Csp and of the titin I27 domain are indicated in figure 6.13. As in the case of 20,80 and 200pN/s the the sum of the lengths of the Csp in each trace was used to estimate the average value of the total step length of the Csp. Averaging the distributions for the Csp results in a mean step length of (19.2 ± 1.0)nm and in a mean unfolding force of (83.2 ± 29.2)pN. The corresponding results from the average of the titin I27 domain are (24.9 ± 0.8)nm and (221.2 ± 34.2)pN.

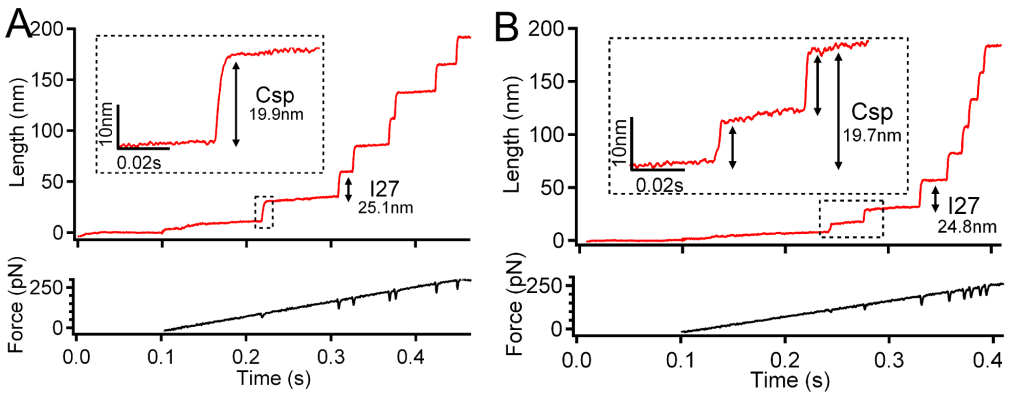

Figure 6.12: Force ramp traces for Csp at a constant force rate of 800pN/s (A) The Csp unfolds via one step (B) The Csp unfolds via two steps

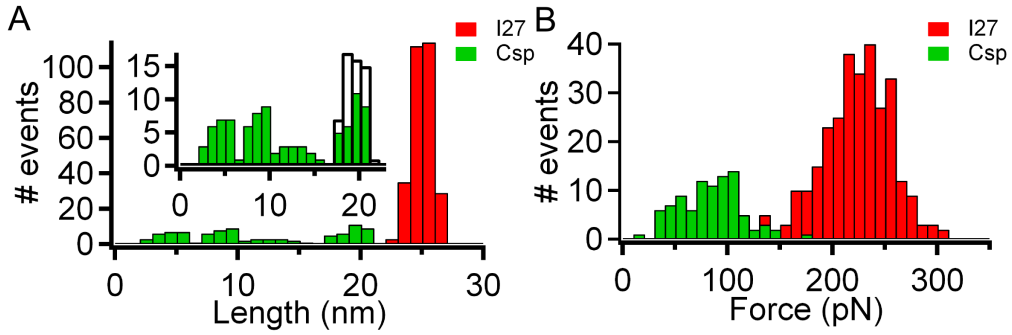

Figure 6.13: Histograms of the distributions of the estimated step length (bin size 1nm) (A) and the measured unfolding force (bin size 10pN) (B) of the Csp and the titin I27 domain at a constant force ramp of 800pN/s

Summary and Discussion

The force ramp measurements reveal a complex mechanical unfolding behavior of the Csp, which confirms the assumption from the constant velocity measurements that the Csp can unfold through intermediate states. Here the application of a precise controlled mechanical force on the Csp identified the distribution of unfolding forces and step lengths with a higher resolution. Table 6.4 summarized the average of the measured force and lengths distributions of the Csp at all for force ramp velocities.

Therein the average value of the sum of step length distributions are listed to estimate the total unfolding length. As it has been observed for the constant velocity experiments the higher the force (increased force ramp rate), the higher the measured unfolding force of the Csp. The unfolding length is around the same value (≈19nm) at all force rates used. The detected step lengths are 4-5 nm smaller than the measured difference in contour lengths from the constant velocity measurements, which is explained by the WLC model, because this model estimates the total length of the polipetide chain at very high force

value approaching infinity, when the extension in x reaches the contour length value L_c.

	Csp			
Force (pN)	58.7 ± 22.2	62.4 ± 24.6	67.8 ± 27.3	83.2 ± 29.2
Length (nm)	19.0 ± 1.1	19.2 ± 1.0	18.8 ± 1.1	19.2 ± 1.0
Force rate (pN/s)	20	80	200	800
Nr. of traces (n)	60	22	63	56
% of traces with several steps	55	55	49	45

Table 6.2: Overview of the average values of the measured force and step lengths for the Csp

Table 6.3 shows the corresponding measured values values for the I27 titin domain.

	I27			
Force (pN)	164.5 ± 22.8	171.4 ± 29.1	200.6 ± 28.0	221.2 ± 34.2
Length (nm)	24.3 ± 0.7	24.5 ± 0.6	24.5 ± 0.7	24.9 ± 0.8
force rate (pN/s)	20	80	200	800

Table 6.3: Overview of the average values of the measured force and step lengths for the I27 titin domain

The probability for the Csp to unfold via several steps has been found to be in the same range under all force ramp condition and lies around 50%. Therefore the detected unfolding behavior of the Csp shows much more diversity than e.g. the one of the I27, which unfolds under all conditions via a single step. Figure 6.14 summarized the described behavior. Therein the length distributions of the Csp were normalized, because of the different amount of traces (n) recorded at each force ramp rates. The normalized distributions reveal that the size of step lengths of Csp can have almost all values between 2 and 20nm (see panel A), which is its total length. Additionally to this complexity, no relation have been found between the unfolding force and the step length, which suggests that the unfolding process of Csp does no occur in a consecutive and force dependent order. Instead small and large parts of the Csp can unfold first, independent of the amount of force (see panel B).

As it has been mentioned before, the probability for the Csp to unfold via one and via multiple steps was found to be equal for all force ramp rates used. Surprisingly, even at a rate of 800pN/s, still clear intermediates were detectable as it has been shown in figure 6.12 panel B. Figure 6.15 summarizes the probability of detecting intermediates during the mechanical unfolding of Csp. In this figure, each bar represents the total unfolding length of Csp in one trace at the corresponding force ramp rate. Therein the color code

from purple to red indicates the amount of single steps counted for Csp in each single trace. E.g. an entire purple bar corresponds to an unfolding of Csp via one step. Every additional color corresponds to an additional step in the trace.

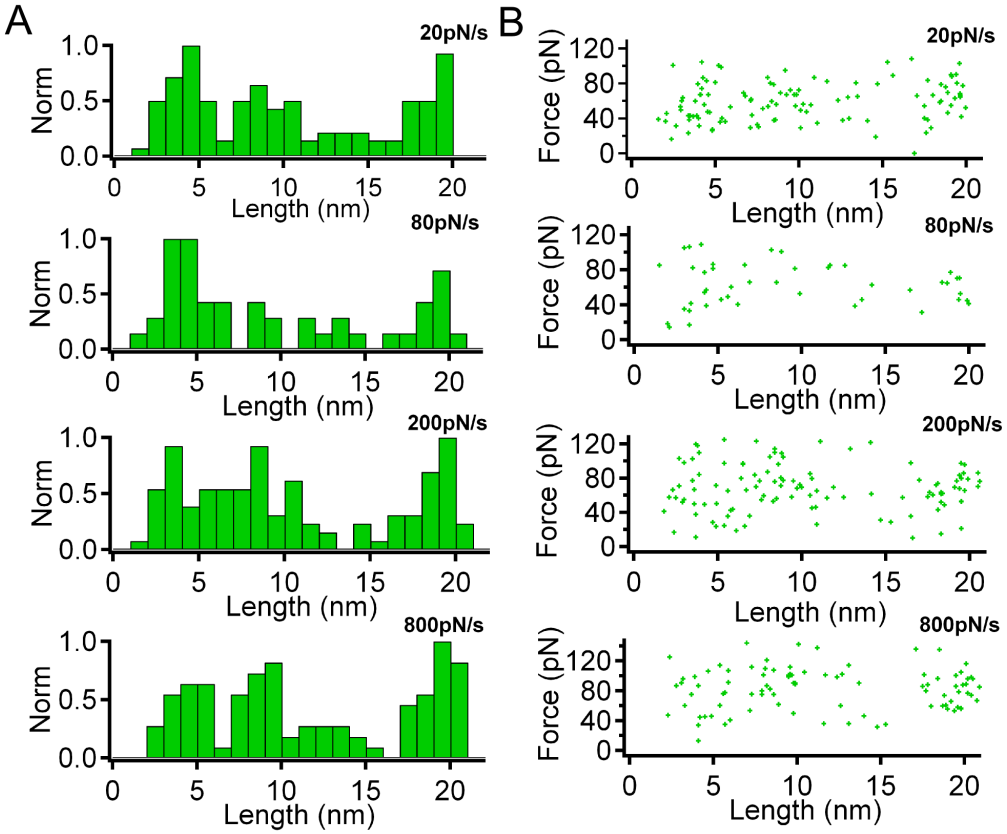

Figure 6.14: (A) Normalized histograms of the distributions of the measured step lengths of Csp for all force ramp rates (bin size 1nm)(B) Corresponding force vs step lengths distributions of Csp for all force ramp rates

Figure 6.15 visualizes that the probability for the Csp to unfold through one or more steps is basically equal at all force ramp velocities. A few traces were detected in which the Csp unfolds through up to 5 steps (see for example figure 6.10 (B)). Table 6.4 gives an overview of the detected population of detected unfolding steps of the Csp and its corresponding percentage values.

The force ramp measurements enabled a close description of the mechanical unfolding behavior of the Csp. However in order to have a higher resolution of the intermediate states during the unfolding of the Csp, force clamp measurements were the experiments of choice as they enable to hold the force precisely constant with time.

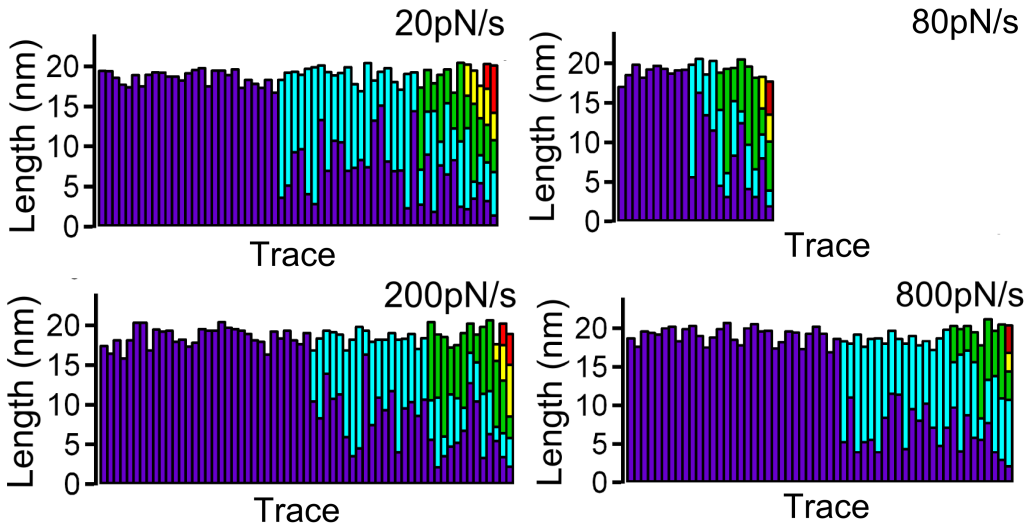

Figure 6.15: Probability of the unfolding pathways of Csp for all force ramp velocities used.

Nr of steps	20pN/s	80pN/s	200pN/s	800pN/s
1	27 (45%)	10 (45%)	32 (51%)	31 (55%)
2	21 (35%)	4 (18%)	18 (28%)	16 (29%)
3	7 (12%)	6 (27%)	10 (16%)	8 (14%)
4	3 (5%)	1 (5%)	1 (2%)	
5	2 (3%)	1 (5%)	2 (3%)	1 (2%)
Nr. of traces	n=60 (100%)	n=22 (100%)	n=63 (100%)	n=56 (100%)

Table 6.4: Populations of different Csp unfolding pathways

6.1.3 Constant force measurements

In the conducted force clamp measurements, the Csp was held under the force of choice for 10 to 15 seconds until the force was switched to 150-170pN in order to reveal the titin I27 unfolding pattern at the end of the trajectory. As in the case of the force ramp measurements, at the beginning of the trace the cantilever was pushed first with $|\tilde{F}| = 1$nN against the surface (pushing). This force value was then decreased within 0.1s to $|\tilde{F}| =$ 100pN, from which the force was then put to the desired force value with the direction from the surface (pulling). Like in the force ramp measurements, in each shown trajectory the red trace represents the length vs time and the black the corresponding force vs time trace. Again the estimated step length of the Csp and of the I27 titin domain are indicated in the figures. Also here the insets show the corresponding unfolding step of the Csp at a higher magnification.

During the force clamp measurements beside the step length also the corresponding dwell

time of the Csp is measured directly. This dwell time is estimated at each unfolding step of the Csp. However just the the dwell time of the step at which the Csp finally unfolds describes the corresponding total unfolding time. The average of the distribution of these total dwell times indicates then the mean first passage time (MFPT) of the Csp.

The measured step lengths of the titin I27 domain at a constant force of 150-170pN will not be mentioned as they were not different from the already measured step lengths from the force ramp measurements. Thus just its mechanical fingerprint is being needed to select the unfolding traces of the Csp for the data analysis.

Again like it was shown in the case of constant velocity and force ramp measurements, figure 6.16 shows two traces as an example of traces which needed to be discarded for the following reasons:

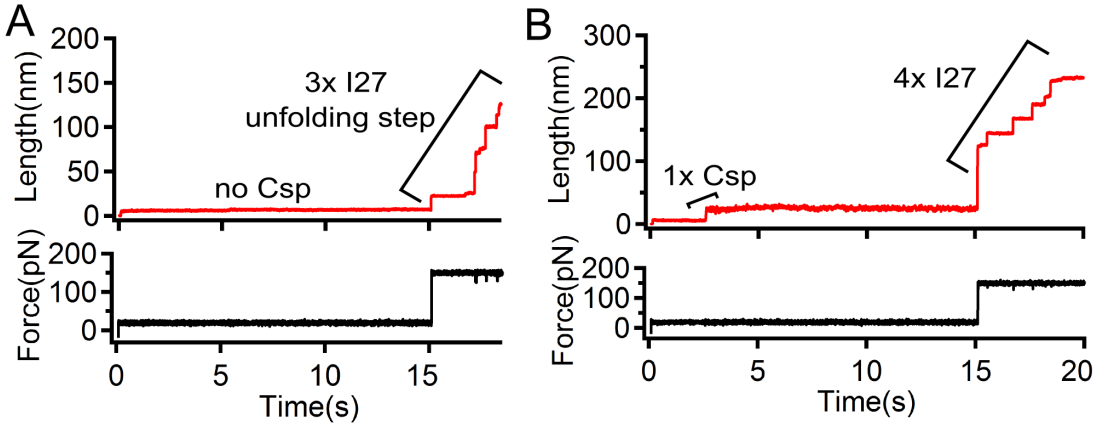

Figure 6.16: Two length and force vs time traces for Csp at a constant force of 20pN as an example of traces which cannot be used for the data analysis (A)-(B).

The first trace (figure 6.16 (A)) indicates no Csp unfolding signal within the first 15s when a force of 20pN is applied. After that several unfolding steps can be detected, but not correlated to the unfolding of the polyprotein domains. In all cases, trace like these cannot be used for the data analysis. In the second trace (figure 6.16 (B)) the Csp seem to unfold at a force of 20pN after around 2.5 seconds, but the fingerprint of the polyprotein construct containing I27 titin domains is not clear. Here just four I27 domains unfolding step can be detected, but the total length is larger than six I27 unfolding steps. This indicates that more than one polyprotein construct has been adsorbed to the cantilever. Therefore similar trace were needed to be discarded from the data analysis.

Constant force measurements at a force of 20pN

Constant force measurements at a force of 20pn reveal a very similar observation of the unfolding of Csp. Two typical force clamp traces at a constant force of 20pN are shown in figure 6.17. At a force of 20pN the Csp can unfold via one unfolding step (panel A). Figure 6.17 (B) shows a trace in which the Csp unfolds via two steps, which confirms the findings from the force ramp experiments.

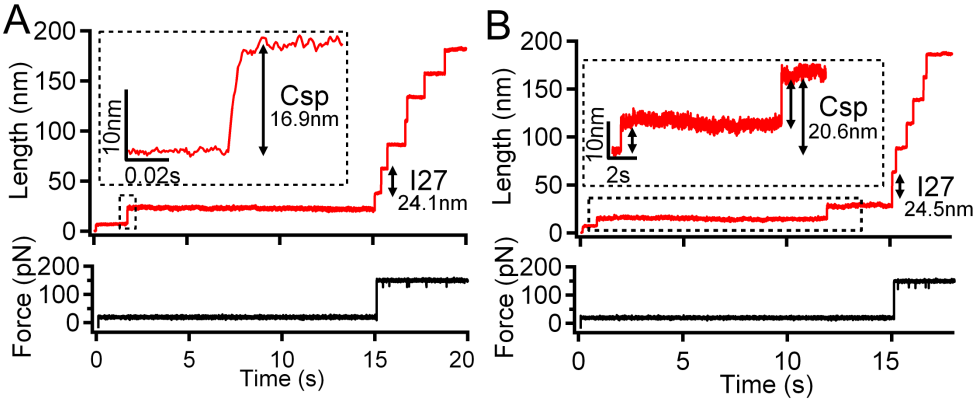

Figure 6.17: Force clamp traces for Csp at a constant force of 20pN (A)Csp unfolds via a single step (B) Csp unfolds via two steps

Additionally, as it can be seen in panel B the Csp can be hold in an intermediate state for a very long time, in this case around 12 seconds. The fluctuations seen in the length vs. time trace can reach ± 2nm (3-4nm amplitude), when using the MLCT cantilever with a spring constant k ≈ 10-20pN/nm. Because of the length fluctuations at 20pN, the measured distribution of the total step length of the Csp is very broad, as depicted in the insets of the shown experimental traces (figure 6.17 (A)-(B)). The lengths measured ranged from 16 to 20nm. Here maybe in some cases the Csp is not fully extended at a constant force of 20pN (see also panel A in figure 6.16).

In total, 52% of all traces (n=56) reveal a single step unfolding behavior, whereas in 48% the Csp unfolds through more than one step, which is a very similar proportion compared to the results of the force ramp measurements.

Figure 6.18 shows the corresponding measured distributions of the step length and total dwell time of the Csp.

The overall distribution of the step lengths (panel A) is widely spread from 2 to 20nm. A peak of lengths around 4nm can be distinguished, which has a similar height than the peak around 15-16nm. Within panel A the white distribution corresponds to the measured total step length from each trace of the Csp. Averaging this sum of the lengths of the

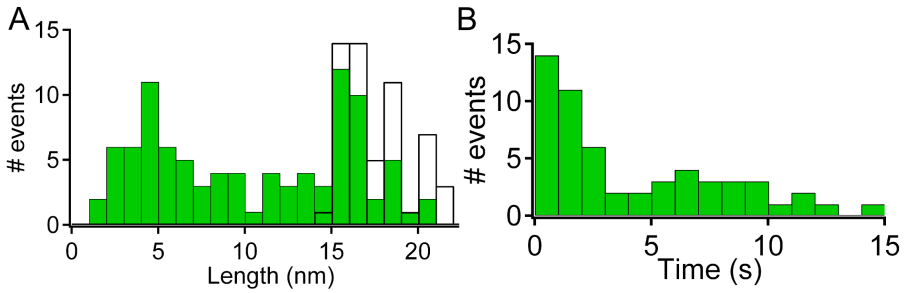

Figure 6.18: Histograms of the distributions of the estimated step length (bin size 1nm) (A) and the measured unfolding total time dwell time (bin size 1s) (B) of the Csp at a constant force of 20pN

Csp results in a mean step length of (17.5 ± 2.0)nm. Panel B shows the distribution of dwell times at which the Csp totally unfolds. Its average resulted in the mean first passage time of (4.1 ± 1.2)s at 20pN. Here the error was estimated using the discussed bootstrap method for the average values, which was done for number of iterations of 100000 and a resampling number of 10 from n=56 data point. Then the standard deviation of the average value was calculated.

Constant force measurements at a force of 40pN

Two examples of force clamp traces at a constant force of 40pN are shown in figure 6.19. Both traces indicate the Csp unfolding signal and five (panel A) or six (panel B) unfolding events of the titin I27 domain. Here again traces are observed in which the Csp unfolds via one (panel A) or via many steps, in the case shown in panel B via four unfolding steps.

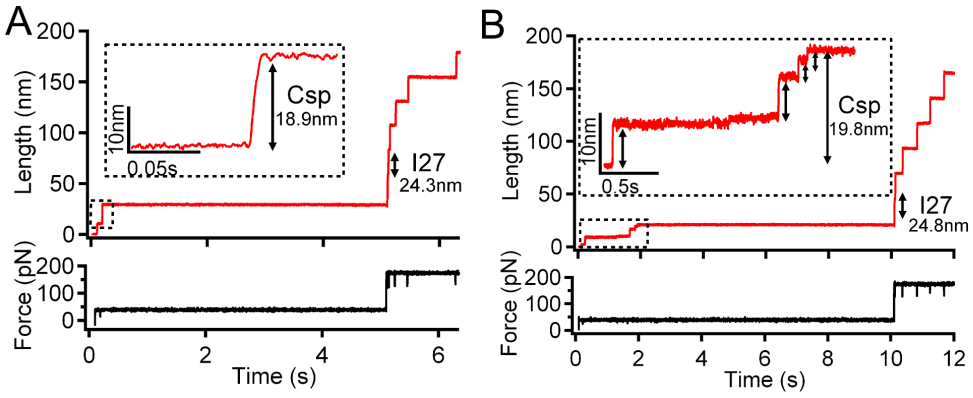

Figure 6.19: Force clamp traces for Csp at a constant force of 40pN (A) Csp unfolds via one step (B) Csp unfolds via four steps

In total around 39% of all traces (n=57) reveal a single step unfolding behavior of the Csp, whereas in 61% of all traces the Csp unfolds via several steps. Comparing to the measurements done at a constant force of 20pN (figure 6.17) a high amount of traces (61%)

indicate a multistep unfolding behavior of the Csp. Even at that force of 40pN dwell times of intermediate states can be found, which reach up to several seconds. Figure 6.19 (panel (B)) indicates such a long dwell time for up to 2s.

Figure 6.20 shows the measured distributions of the step length and total dwell time of the Csp. The overall distribution of the step lengths (panel A) is again widely spread from 2 to 20nm. However in contrast to the corresponding histogram at 20pN (fig 6.18A), here the majority of step lengths measured are between 2 and 7nm. Also the amount of detected small steps is higher than the higher step values around 19nm.

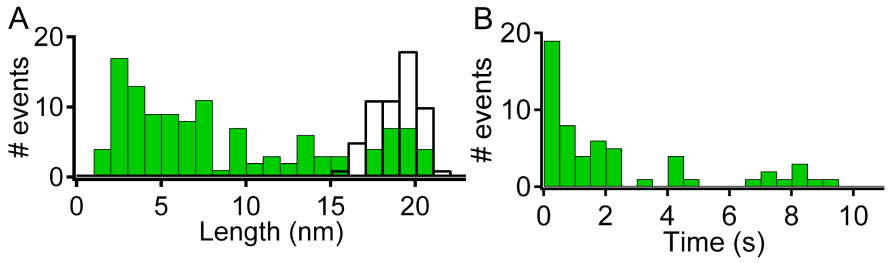

Figure 6.20: Histograms of the distributions of the estimated step length (bin size 1nm)(A) and the measured unfolding total time dwell time (bin size 0.5s)(B) of the Csp at a constant force of 40pN

The average of the sum of the step lengths the Csp (white distribution in panel A) results in a mean step length of (18.8 ± 1.3)nm. The corresponding total dwell time distribution is shown in panel B, which average results in a mean first passage time of (2.3 ± 0.9)s, in which again the error was estimated by applying the bootstrap method.

Constant force measurements at a force of 60pN

At a force of 60pN the Csp unfolding behavior is very similar to the one observed at 40pN (see figure 6.21).

Traces can be detected, in which the Csp again unfolds via one unfolding step (panel A) or multiple steps. Panel B shows an example in which up to five unfolding steps could be counted. Surprisingly, even measurements at 60pN reveal stable measurements of intermediate lifetimes of the Csp. For example dwell times can be found, which reach up to 0.5-1s (panel B). Around 39% of all traces (n=56) reveal a single step unfolding behavior of the Csp, whereas in 61% Csp unfolds via several steps, which is the same proportion as it had been detected before at 40pN.

Figure 6.22 shows the measured distributions of the step length and total dwell time of the Csp.

The overall distribution of the step lengths (panel A) is again widely spread from 2 to

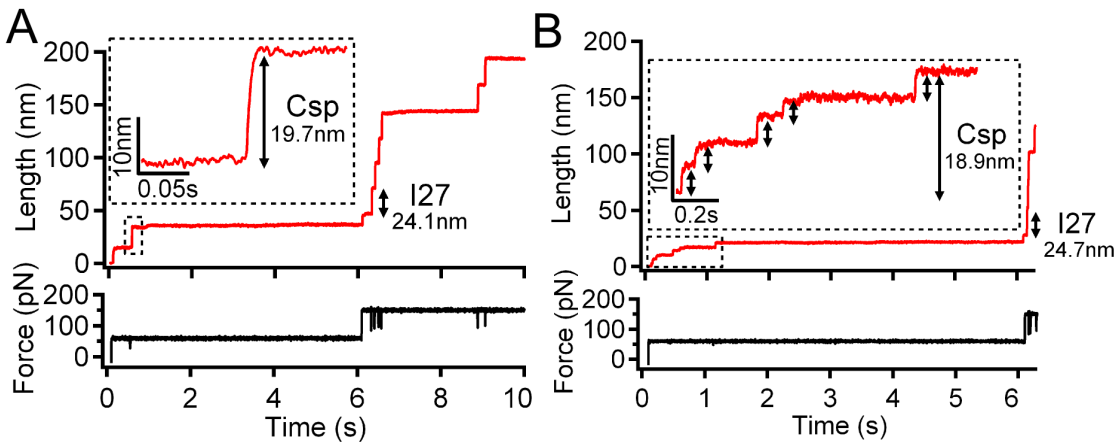

Figure 6.21: Force clamp traces for Csp at a constant force of 60pN (A) Csp unfolds via one single step (B) Csp unfolds via five steps

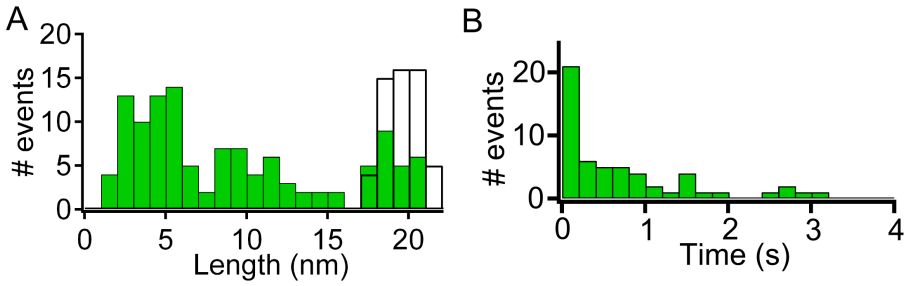

Figure 6.22: Histograms of the distributions of the estimated step length (bin size 1nm) (A) and the measured unfolding total time dwell time (bin size 0.2s)(B) of the Csp at a constant force of 60pN

20nm. Like it has been observed at a consatnt force of 40pN, at 60pN the majority of step lengths measured are between 2 and 7nm and also between 8 and 11nm. The average of the sum of the step lengths the Csp (white distribution in panel A) results in a mean step length of (19.6 ± 1.1)nm. The corresponding total dwell time distribution is shown in panel B, which average gives a in mean first passage time of (0.8 ± 0.3)s.

Constant force measurements at a force of 80pN

At a force of 80pN the Csp unfolds almost every time via one step, as it can be seen in panel A in figure 6.23.

However, also at that high force traces are observed in which the Csp unfolds via two clear steps (figure 6.23 panel B). In sum 70% of all traces (n=57) reveal a single step unfolding behavior of the Csp, whereas in 30% Csp unfolds via two steps. However the fact that in the majority of traces the Csp is detected to unfold via one step, could be due to the time resolution limit of the force clamp AFM. Figure 6.24 shows the measured distributions of the step length and total dwell time of the Csp.

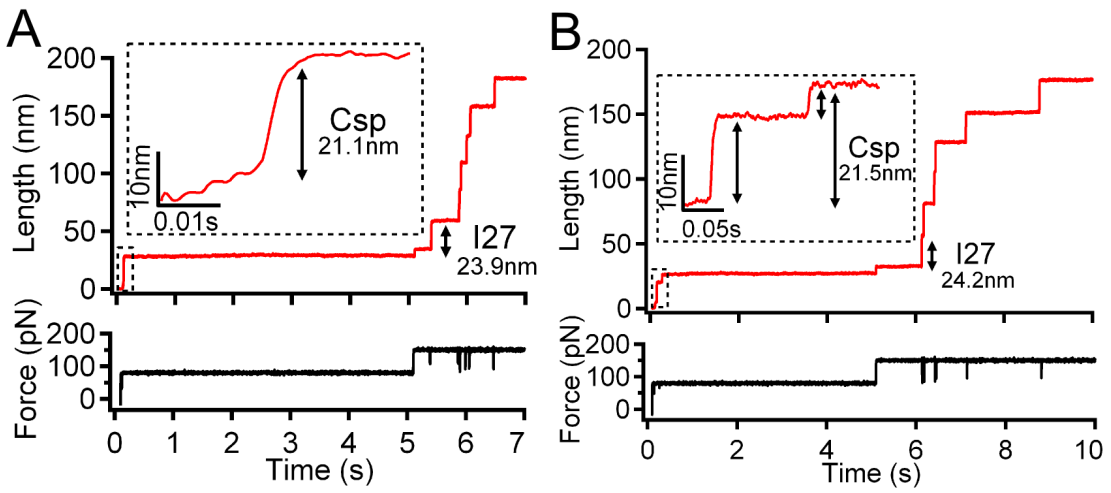

Figure 6.23: Force clamp traces for Csp at a constant force of 80pN (A) Csp unfolds via one step (B) Csp unfolds via two step

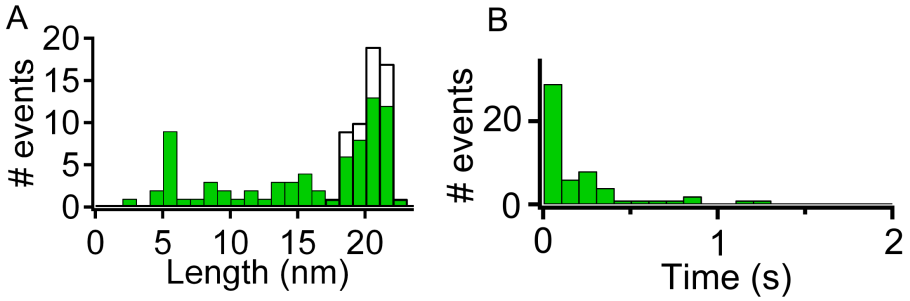

Figure 6.24: Histograms of the distributions of the estimated step length (bin size 1nm) (A) and the measured unfolding total time dwell time (bin size 0.1s) (B) of the Csp at a constant force of 80pN

The overall distribution of the step lengths (figure 6.24A) is much less spread compared to the corresponding distributions at smaller forces. The main peak is the total length of the unfolded Csp, which is around 20nm. However a small peak still exists around 5nm. The average of the sum of the step lengths the Csp (white distribution in panel A) results in a mean step length of (20.3 ± 1.1)nm. The corresponding mean first passage time was estimated to be (0.4 ± 0.2)s, derived from the corresponding total dwell time distribution shown in panel B.

Constant force measurements using a force sequence from 20 to 80pN

In order to corroborate the discovered complex Csp unfolding behavior at a constant force, experiments were conducted in which a force sequence was applied to one and the same $I27_3$-Csp-$I27_3$ single polyprotein molecule. Within this experiments a sequence of force pulses (20, 40, 60 and 80pN) was applied to the polyprotein construct for 10s each. After each pulse the force was released to zero for another 10s, in order to allow the Csp to

refold, before the next force pulse was applied. In the end of the trace a force pulse of 150pN was applied in order to detect the mechanical fingerprint of the titin I27 domains. With this experimental protocol the unfolding of the Csp is fully separated from the one of the titin I27 domains, as the I27 domain rarely unfolds at low force from 20 to 80pN within 10s. Figure 6.25 shows two of this force sequence containing trajectories on the Csp in panel A and B.

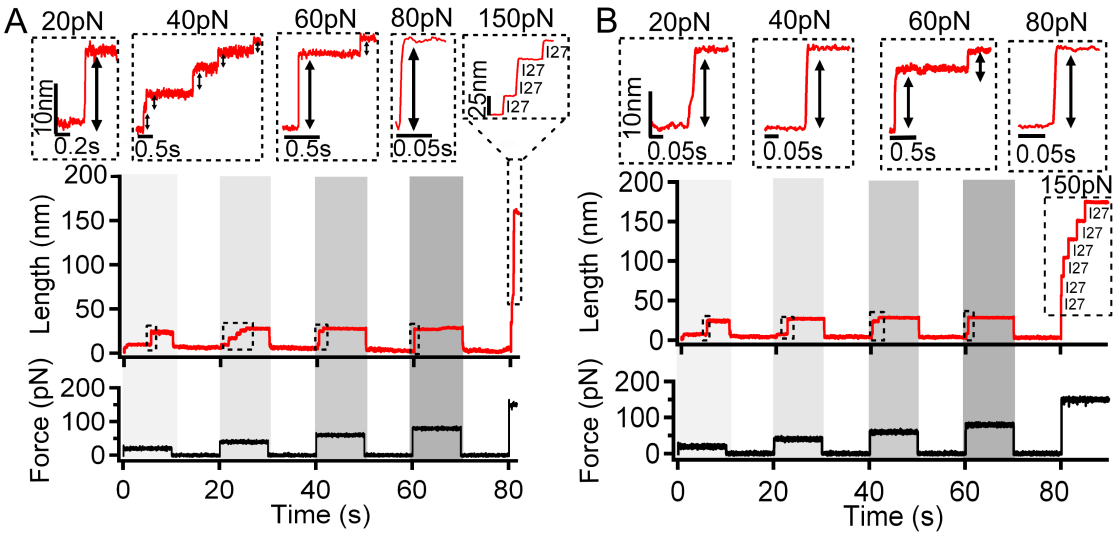

Figure 6.25: Multiple force sequence trajectories with Csp

Both trajectories have the important I27 fingerprint at the end of the trace where the force pulse of 150pN is applied. This validates the reliability of the trace. In panel A four and in panel B six I27 domains could be distinguished. The in gradient grey colored bars indicate the time of the applied force pulse from 20pN (less grey) to 80pN (dark grey). The four insets at the top the each trajectory magnify the unfolding step(s) of the Csp at the corresponding force. The last inset at 150pN demonstrates the molecular fingerprint of the titin I27.

The main advantage of this type of experiment is that the different forces are applied at one and the same Csp, whereas in the measurements conducted before the mechanical forces were applied on different individual molecules of the same sample. I.e this kind of measurement has a real single molecule character.

In sum, figure 6.25 reveals the complex unfolding behavior of Csp under a constant force from 20 to 80pN in a single trace. In panel A the Csp unfolds through one, five and two steps, whereas in panel B the Csp unfolds through one and two steps.

Surprisingly, the protein construct could be held on the cantilever for such a long time (up to 90s), although the cantilever was approaching and retracting the surface five times

and no specific binding was used for the attachment of the protein sample. However these measurements showed that in order to discover one trajectory, in which the Csp unfolds during all applied force pulses and additionally at least four titin I27 domains unfold at the end of the trace, many experiments needed to be conducted. This easily resulted in a positive pick up rate as low as around 0.01%.

Summary and Discussion

The force clamp measurements reveal the mechanical unfolding behavior of the Csp, which confirms that the Csp can unfold through stable intermediate states and this under a constant mechanical force ranging from 20 to 80pN. The following table 6.5 summarized the average of the measured total unfolding length and the estimated MFPT of the Csp from the single force pulse experiments conducted before.

	Csp			
Length (nm)	17.5 ± 2.0	18.8 ± 1.3	19.6 ± 1.1	20.3 ± 1.1
MFPT (s)	4.1 ± 1.2	2.3 ± 0.9	0.8 ± 0.3	0.4 ± 0.2
Force (pN)	20	40	60	80
Nr. of traces (n)	56	57	56	57
% of traces with several steps	48	61	61	30

Table 6.5: Overview of the average values of the measured unfolding length and mean first passage time (MFPT) of Csp

Here the higher the force value, the larger the measured total unfolding length of the Csp, which is in accordance to the WLC model. The average of the unfolding length increases from 17.5nm at 20pN to 21nm at 80pN. Additionally, they are in the range of the measured total unfolding lengths in the force ramp experiments (see table 6.4).
The estimated mean first passage times allow to derive a unfolding rate k_u^0 and a distance to the unfolded state Δx_u of the Csp at zero force using the described Bell formula (see chapter 2.4 and equation 2.3). From the MFPT values, the corresponding mean unfolding rates are obtained and plotted in a semi-log graph vs the applied constant force (see figure 6.26).
The linear regression of the unfolding probability P_{unf} results in an k_u^0 value of $0.07s^{-1}$ and a Δx_u value of 0.20nm. Here both values were obtained from the two-state model from Bell in order to have a crude approximation, although the force ramp and force clamp measurements reveal a clear multistate unfolding behavior of the Csp.

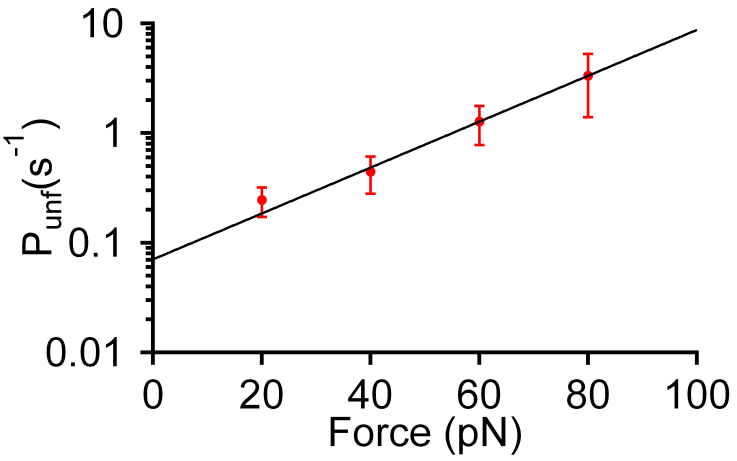

Figure 6.26: The semi-log plot of the mean unfolding rate (unfolding probability P_{unf}) vs the force measured for the Csp

In contrast to the force ramp measurements, the probability for Csp to unfold via one and multiple steps was found to differ depending on the applied force within the force clamp experiments. Whereas in the majority of cases at 20 and 80pN Csp unfolds via one or two steps, in the case of 40 and 60pN the probability is higher for Csp to unfold via multiple steps. Figure 6.27 summarizes the probability of detected intermediates during the mechanical unfolding of Csp, plotting again the total length represented by a single bar in the same color code from purple to red as it has been done for the force ramp measurements in figure 6.15. Here it can be seen that the amount of constant force modulates the unfolding pathway of the Csp.

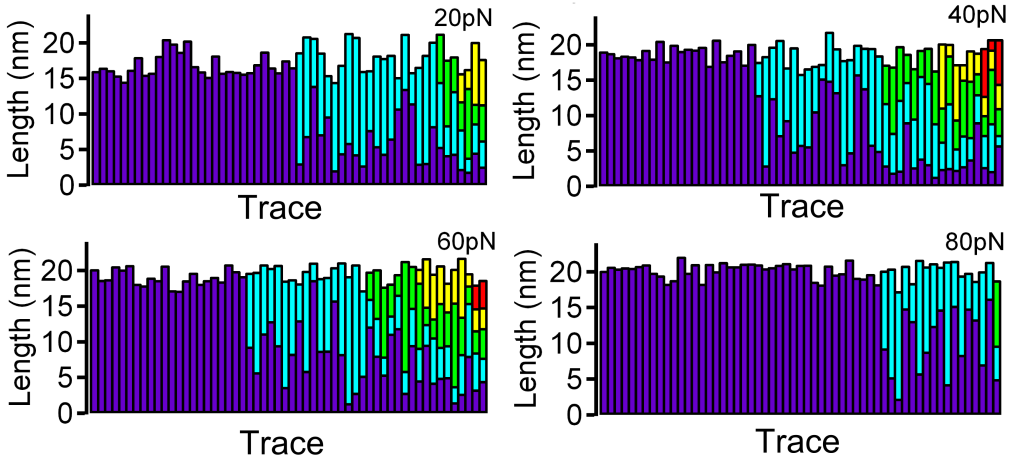

Figure 6.27: Probability of the unfolding pathways of Csp for all constant force values used.

At 40 and 60pN a few traces were detected in which the Csp unfolds through up to 5 steps (see also figure 6.21 (B)). Table 6.6 represents the number of traces of unfolding steps from one to five in relation to the corresponding amount of force.

Nr of steps	20pN	40pN	60pN	80pN
1	29 (52%)	22(39%)	22 (39%)	40 (70%)
2	20 (36%)	18(32%)	17 (30%)	16(28%)
3	3 (5%)	8(14%)	7(13%)	1 (2%)
4	4 (7%)	6 (11%)	8 (14%)	
5		3(4%)	2(4%)	
	n=56	n=57	n=56	n=57

Table 6.6: Populations of different Csp unfolding pathways

Apart from the probability of Csp to unfold through different pathways, in fact neither a consecutive order of unfolding steps nor a repeating pattern of specific step lengths could be distinguished. Figure 6.28 sums the step size distribution of all measured step lengths of Csp at all four constant forces together.

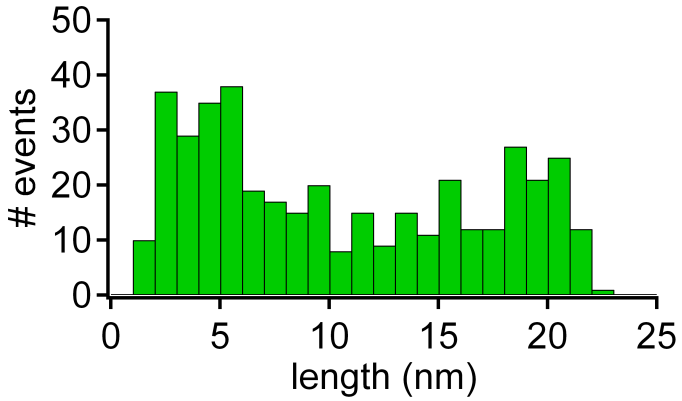

Figure 6.28: Distribution of unfolding step lengths of Csp for all four constant forces 20-80pN.

Herein several peaks can be identified from 2-5nm, 9nm, 11-13nm, 15nm and 18-20nm. However because the distribution depends always on the chosen bin size (e.g. 1nm bin size in figure 6.28) and the Csp length histograms are very broad suggests that a conventional histogram analysis is not sufficient enough to resolve the underlying complexity. Therefore as an alternative a k-means cluster analysis has been conducted in order to reveal further insights. The analysis revealed that at each constant force the length data could be perfectly distributed into six clusters. Furthermore the properties of each cluster looked very similar instead of the last cluster which represents the extended total unfolding length. Because the cluster properties looked very similar at each force a global cluster analysis was performed with all step lengths together from all the four constant force measurements. Here it was found that eight clusters fits the distribution of step lengths best These clusters

counted from C1 to C8 are listed in the figure 6.29 with the corresponding ±std error.

C1 (2.6 ± 0.5)nm C5 (12.6 ± 0.9)nm
C2 (4.6 ± 0.6)nm C6 (15.6 ± 0.6)nm
C3 (6.7 ± 0.7)nm C7 (18.4 ± 0.6)nm
C4 (9.3 ± 0.7)nm C8 (20.6 ± 0.6)nm

Figure 6.29: Distribution of unfolding step lengths of Csp for all four constant forces 20-80pN.

The three longest lengths C6-C8 correspond to the total unfolding length of the Csp unfolding state from 20pN to 80pN. The population of each of the eight clusters was then estimated not just for all forces together like in the distribution shown in figure 6.28, but also for each force separately. Furthermore also the transition matrix from one cluster to another was estimated from the experimental data input. All in all the eight clusters reveal a stepwise increase of a defined length (≈ 2.5nm) comparable with the peaks found in the length histogram in figure 6.28.
Figure 6.30 visualizes the results of the cluster analysis and the corresponding transition matrix together for all four forces.

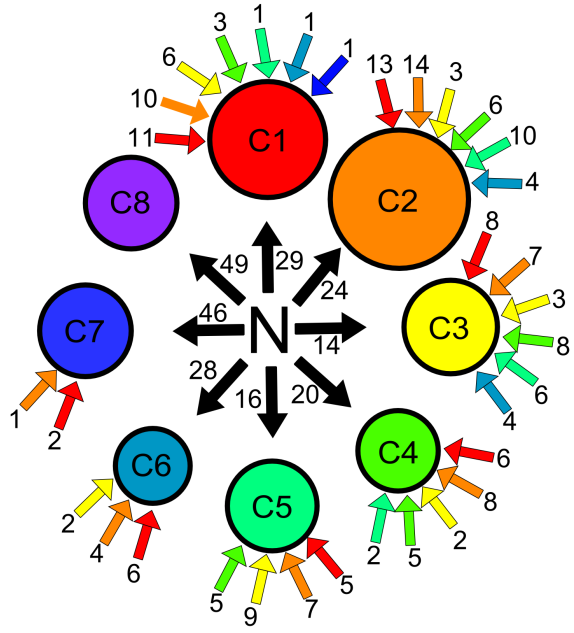

Figure 6.30: Representation of the cluster populations and the corresponding transition matrix of the distribution of step lengths of Csp for all constant forces.

Therein each cluster from figure 6.29 is represented with a circle, which diameter corresponds to the estimated population of the cluster. The transitions either from the native state or from one cluster to another are visualized by arrows which point at the cluster. The numbers indicate the population of each transition and the color of each arrow indi-

cates the the cluster from which the transition starts. Here also a transition is possible from one cluster to the same cluster, for example when during unfolding of the Csp one step length is followed by a step length of the same cluster. The huge population of clusters C1 and C2 and the corresponding transitions represents the detected amount of multiple step unfolding of Csp during all four constant force measurements.

Starting from the cluster in figure 6.29, the same representation used in the figure 6.30 above was now applied to the step length distributions measured for each single force (see figure 6.31).

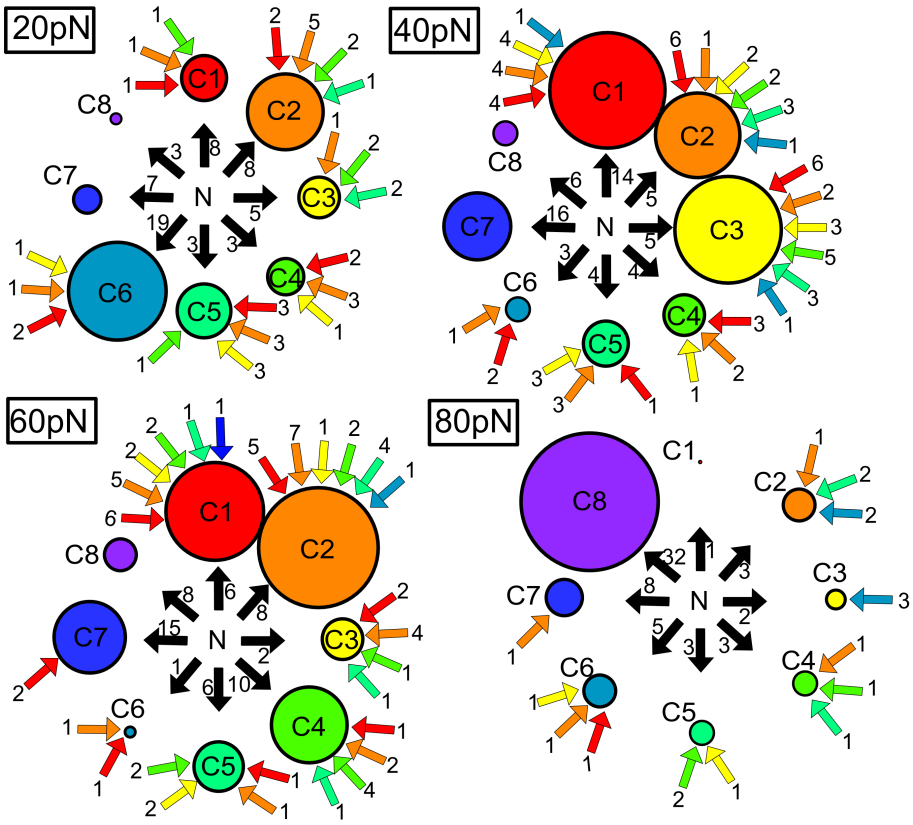

Figure 6.31: Representation of the cluster populations and the corresponding transition matrix of the distribution of step lengths of Csp separated by each constant force used from 20 to 80pN.

This figure visualizes the differences in the unfolding pathways of Csp detected especially between 20pN and 40pN-60pN. Here the already described variability for the Csp to unfold via different unfolding pathways at a force of 40 and 60pN (see e.g. figure 6.15) is represented by the high population of clusters C1, C2 and C3 and the high amount of transition into these clusters. The cluster with the largest unfolding length changes from C6 to C8, while increasing the force from 20 to 80pN, as a higher force increases the probability for Csp to be stretched to a larger end-to-end distance. At 80pN almost all transitions (32) occur directly from the native state N to the cluster C8.

However a consecutive repeating pattern like a certain unfolding sequence of step lengths of Csp at a constant force between 20 and 80pN cannot be figured out. Furthermore when changing the force from 20 to 40 or 60pN, the unfolding behavior of Csp changes, showing a even higher variability in unfolding pathways.

In sum the conducted force clamp measurements revealed the high complexity of the unfolding behavior of Csp under a mechanical force. Additionally it has been shown that the amount of force applied to the Csp can directly effect its unfolding pattern. In order to investigate if the experimentally measured unfolding step lengths could be referred to the structural elements like the five beta strands β_1-β_5 and the loop of the Csp (see figure 3.3) SMD simulations have been conducted and its results will be shown in the next section. However, before the SMD simulations will be discussed, first the stochastic kinetic simulations will be introduced showing their capacity of reproducing the experimental results of Csp in force ramp and force clamp.

6.2 Stochastic kinetic simulations on the Csp

Stochastic kinetic simulation were performed on the Csp for both cases, for a force changing linearly with time (force ramp) and for a constant force. As the simulation only implies a model describing two states, the folded and the extended unfolded state, it was only compared to corresponding force ramp and constant force experimental traces in which just a single step unfolding was observed for the Csp.

The amount of simulated traces was then equal to the number of experimental traces showing one step unfolding under the corresponding experimental condition.

The result of the simulation mainly depends on four variables describing the protein stability under zero mechanical force. These are the unfolding (k_u^0) and the folding rate k_f^0, the distance to the unfolded state Δx_u and the approximate length of the unfolded state x_u^0 of the Csp, although the last parameter is just influencing the final extended length of the protein and not its stability. Here parameters from the literature can be found, which state an unfolding rate of $k_u^0 \approx 0.02s^{-1}$ and a folding rate of $k_f^0 \approx 550s^{-1}$ for the Csp estimated in bulk [186]. The value found for the Δx_u was 0.49nm, which was estimated by constant velocity AFM measurements before [128]. As there is no existing value of the dimension of the x_u value, here a value of 1nm was being considered as a staring point.

The simulations were conducted at a force ramp rate of 20pN/s and at a constant force of 40pN with a spring constant $k_c \approx 15pN/nm$, which is the average value normally obtained for the experimentally used MLCT cantilever D (k $\approx 10-20$pN/nm). The simulations were

performed with a sampling rate of 1ms, which was introduced for filtering the simulated data with a timestep of 0.5ms.

6.2.1 Stochastic kinetic simulations of the force ramp measurements at a rate of 20pN/s

The simulations were conducted under two different conditions. The first simulation (called simulation I) was using the before mentioned values from the literature ($k_f^0 \approx 550s^{-1}$, $k_u^0 \approx 0.02s^{-1}$ and $\Delta x_u \approx 0.49nm$ [128, 186]). The second simulation (called simulation II) was using the experimental estimated values from the constant force MFPT data of the Csp. Therein the values $k_f^0 \approx 550s^{-1}$, $k_u^0 \approx 0.07s^{-1}$ and $\Delta x_u \approx 0.2nm$ were taken. As a comparison the experimental data from all one step Csp unfolding trajectories at a force ramp rate of 20pN/s has been used (see figure 6.6 A). Here all the n=27 experimental traces were taken and therefore n=27 traces were produced in both simulations in order to have the same amount of traces and therefore to have comparable histogram distributions. Figure 6.32 compares the corresponding histogram distributions from the measured length and unfolding force of the experimental data and the two simulations under different conditions.

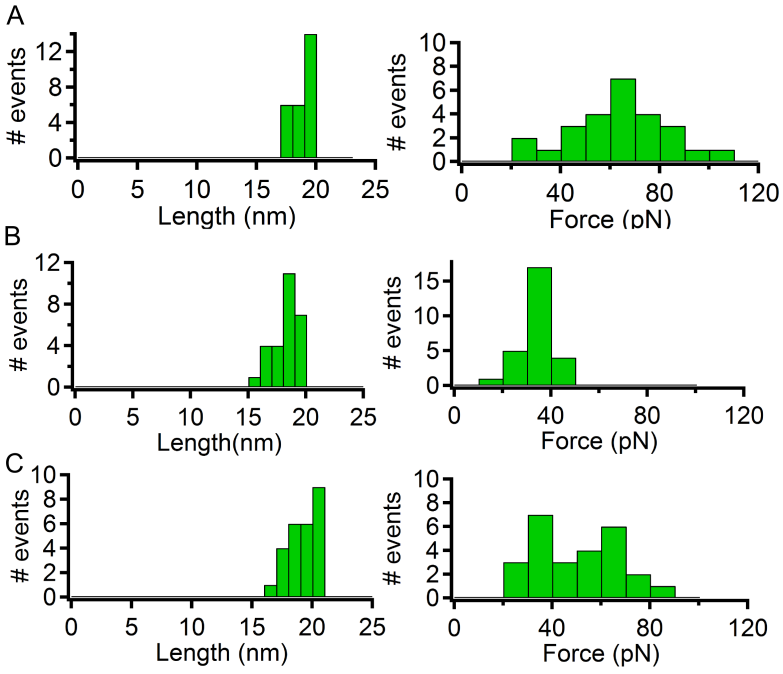

Figure 6.32: Histogram of the distributions of the unfolding length (bin size 1nm) and unfolding force (bin size 10pN) comparing the experimental results at a force rate of 20pN/s (A) with the stochastic kinetic simulations (B)-(C). (B) The stochastic simulation with the values from the literature (simulation I) (C) The stochastic kinetic simulation with the experimental estimated MFPT values (simulation II)

Table 6.8 compares the estimated average values obtained from the distributions of length and unfolding force from figure 6.32.

	Experimental data	Simulation I	Simulation II
Length (nm)	18.8 ± 0.8	18.2 ± 1.1	19.1 ± 1.2
Force (pN)	64.3 ± 19.9	33.7 ± 6.3	50.8 ± 17.4
Nr of traces	27	27	27

Table 6.7: Populations of different Csp unfolding pathways

The results show that clearly the simulation II with the experimentally obtained values fits the experimental data at 20pN/s better than the simulation I. Compared to the experimental estimated average value (58.7 ± 22.2)pN of the unfolding force at 20pN/s (see table 6.4), which includes also the unfolding forces measured at all multiple step lengths, the corresponding results from simulation II fit very well.
Figure 6.33 shows the comparison of an experimental one step Csp unfolding trace at a force ramp of 20pN/s with a simulated trace (simulation II).

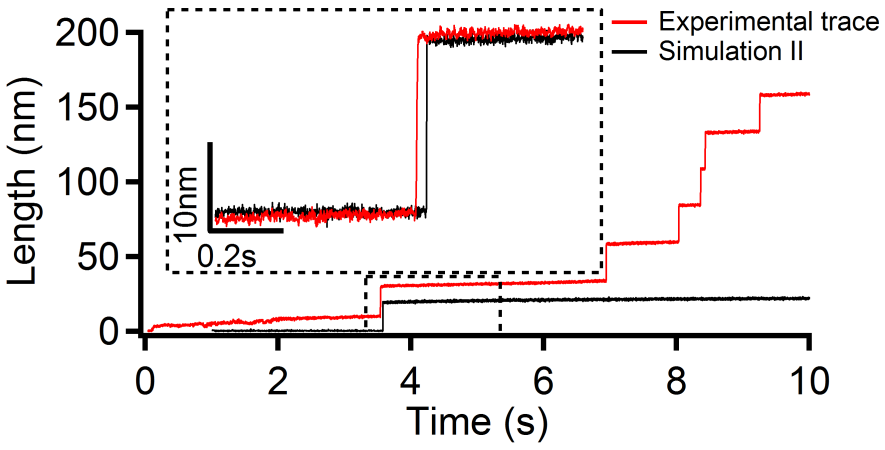

Figure 6.33: Comparison of an experimental trace at 20pN/s (red) with a simulated traces (black, simulation II)

The simulated trace in figure 6.33 shows very a good agreement with one selected experimental trace. In the inset of figure 6.33 both traces were overlapped for a better comparison, although there is a difference in height between both traces (see whole trajectory in figure 6.33). The existing difference in height between both traces can be explained because the experimental trace is not taken from a single Csp protein. Moreover the experimental trace is taken on a polyprotein construct and its initial extension before the Csp unfolds must be taken into account, which is missing in the simulations. Furthermore

the experimental trace starts at a force of -20pN, whereas the stimulation starts directly at 0pN. Therefore the simulation start in figure 6.33 at 1s to be in the same time scale like the experimental trace. However the slope and the fluctuations of the simulated trace are in very good agreement with the experimental trace. The fluctuations of the simulated trace where adjusted to the experimental one by using a sampling time (external filter) of 1ms. Also it has to be considered that the experimental trace has an additional slope in time, which corresponds to a linear increase of the total length of the polyproteinchain as the force also increases linearly. This was not included into the simulation, therefore the slope is slightly different between the experimental and the simulated trace.
All in all the force ramp experimental data of the Csp could be well reproduced by the stochastic kinetic 2-state simulations when using the experimental derived values for the unfolding rate $k_u^0 \approx 0.07s^{-1}$ and distance to the unfolding state $\Delta x_u \approx 0.2nm$ and a sampling time of 1ms. In the following section constant force simulations were conducted under the same two conditions and compared to the corresponding experimental data.

6.2.2 Stochastic kinetic simulations of the constant force measurements at 20pN

Here again the first simulation (simulation I) was using the values from the literature [128, 186] and the second simulation (simulation II) the experimental estimated values from the constant force MFPT data. For the comparison the experimental data from all one step Csp unfolding data at a force of 20pN has been used (see figure 6.17 A). In this case n=29 experimental traces were selected and therefore n=29 trajectories were calculated in both simulations.
Figure 6.34 compares the corresponding histogram distributions from the measured length and dwell time of the experimental data and the two simulations under different conditions. The estimated average values obtained from the distributions of length and dwell time force from figure 6.32 are compared in table 6.8.
Both simulations show a lower average value from the unfolding length distribution compared to the experimental results, which values is around (16.9±1.6)nm when counting just the 29 experimental traces which show 1 step unfolding of Csp. This slightly discrepancy should come from the experimental conditions as a constant force of 20pN is at the limit of resolution when using the MLCT cantilever the force (deflection) of the cantilever can fluctuate around ± 3pN. As the force of 20pn is still comparable low a small change in force could change the measured extended length of the Csp by 1-2nm. The dwell time

distribution shows differences for the second simulation (simulation II) compared to the experimental results. Here the MFPT value estimated in simulation I is very close and the one from simulation II to large compared to the experimental value. However when compared to the real experimental values including the multistep unfolding behavior of Csp, the MFPT from simulation II is in very good agreement with the experimental values (4.1 ± 1.2)s (see table 6.6).

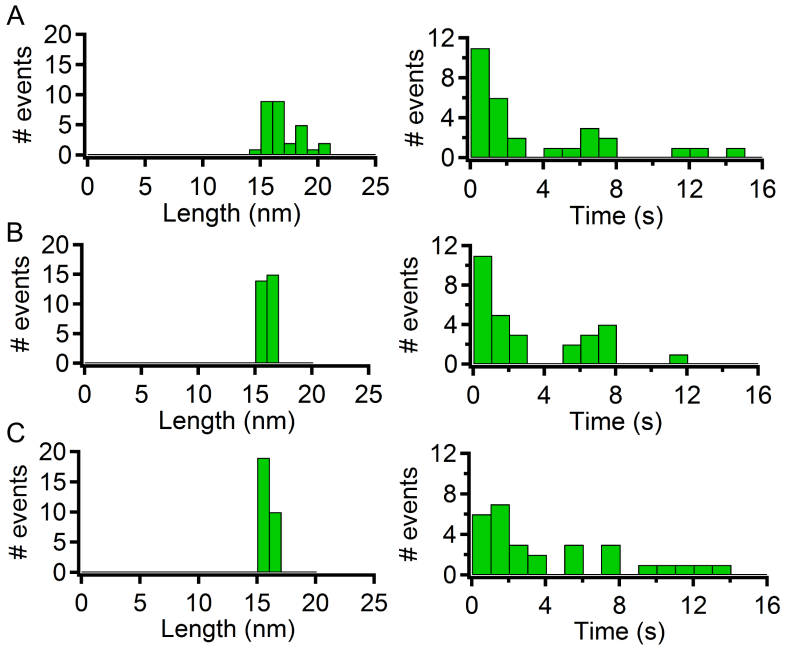

Figure 6.34: Histogram of the distributions of the unfolding length (bin size 1nm) and dwell time (bin size 1s) comparing the experimental results at a constant force of 20pN (A) with the stochastic kinetic simulations (B)-(C). (B) The stochastic simulation with the values from the literature (simulation I) (C) The stochastic kinetic simulation with the experimental estimated MFPT values (simulation II)

	Experimental data	Simulation I	Simulation II
Length (nm)	16.9 ± 1.6	15.9 ± 0.4	15.8 ± 0.3
Dwell Time (s)	3.4 ± 1.3	3.1 ± 1.0	4.3 ± 1.3
Nr of traces	29	29	29

Table 6.8: Populations of different Csp unfolding pathways

The comparison of an experimental one step Csp unfolding trace at a constant force of 20pN with one simulated trace (simulation II) is depicted in figure 6.35. The simulated trace in figure 6.35 shows a very good agreement with the selected experimental trace. In the inset both traces were again overlapped for a better comparison, although there is a difference in height which separate both traces (see real trajectory). As detected before in

the force ramp simulation, the difference in height in both traces can be explained because the initial extension of the polyprotein must be taken into account, which is missing in the simulations. The noise of the simulated trace is in very good agreement with the experimental trace. In this simulation again a time filter of 1ms was used to have a better comparison to the noise of the experimental trace.

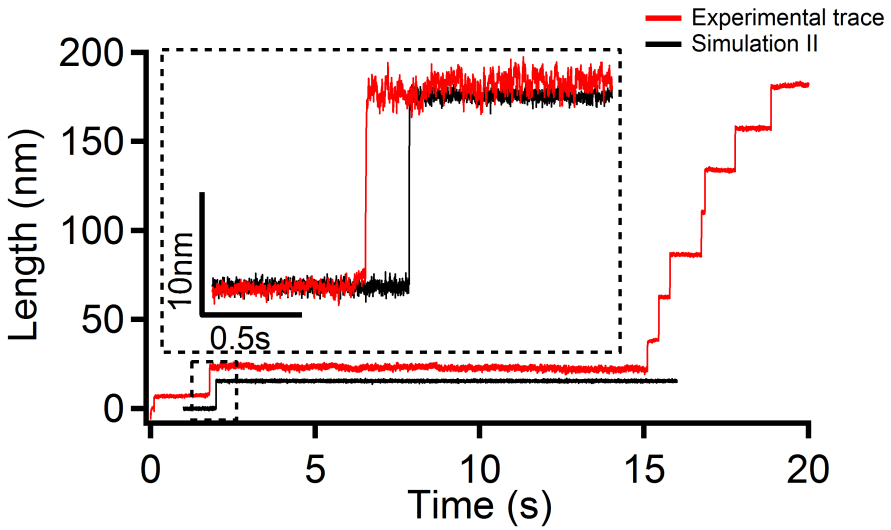

Figure 6.35: Comparison of an experimental trace at 20pN (red) with a simulated traces (black, simulation II)

All in all the constant force experimental data of the Csp could be well reproduced by the stochastic kinetic 2-state simulations when using the experimental derived values for the unfolding rate $k_u^0 \approx 0.07s^{-1}$, a distance to the unfolding state $\Delta x_u \approx 0.2nm$ and a time filter of 1ms, when the real calculation time is 0.5ms.

However here it becomes evident that the simulation II results would fit the overall experimental data of the Csp including its multiple steps better, than when just using the single step traces. This is related to the fact that the simulation II was performed with the experimentally derived MFPT values from all experimental data of the Csp using the Bell model. For the same reason the distributions in length and time from the simulation I, which used the input values from the literature, does not fit the experimental data that good, because of the different values of k_u^0 and Δx_u and because these values were derived from constant velocity measurements, which is a complete different experimental condition compared to force ramp and constant force experiments. During conducted stochastic kinetic simulations at a constant force of 20pN it was observed that the calculated fluctuations of force, which were derived from the free oscillations of the cantilever already affected the Csp folding and unfolding rate under force. Therefore an external damping

factor of 1/3 needed to be introduced in order keep the simulated result close to the experimentally observed result. This damping factor can also be explained by the time of the feedbackloop it takes to track the signal from the photodiode, which was not included into the stochastic kinetic simulations.

6.3 SMD simulations on Csp

All conducted SMD simulations started from the same MD equilibrium simulation. Therefore the first section will provide the MD results. Then the constant velocity and constant force SMD simulations will be shown. The 3Dim structure of the Csp is shown in figure 6.36 indicating especially the five beta strands β_1-β_5, the loop and the N and C terminus to have a better orientation for the following sections.

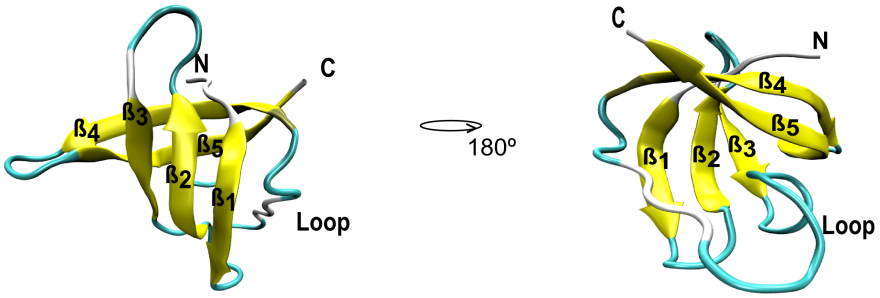

Figure 6.36: The 3Dim structure of the Csp is shown and the structural elements are indicated.

6.3.1 MD equilibrium simulation

Before the MD equilibrium simulation was started, the waterbox containing the Csp (1G6P.pdb) needed to be designed. Finally the waterbox was constructed with the dimension (x y z) of (0 0 0)(30 7 7)nm and contained 141296 atoms. Then the Csp inside the waterbox was minimized using steepest decent for 0.2ns and equilibrated for 1ns.
After the simulation was run, the RMSD value was calculated for each timestep of the equilibrium simulation, which compares the position of the backbone atoms of the original and the simulated structure and which is shown in figure 6.37 A.
Figure 6.37 B shows the corresponding equilibration MD of the turned Csp in order to conduct constant force SMD simulations, in which the force is applied on the C_α atom of the N-terminal residue while the C_α atom of the C-terminal residue is being fixed. As it can be seen the RMSD value is not changing much during the MD simulation and stays below three angstrom (<0.3nm). In the MD shown in figure 6.37 (A) the Csp turns with time a bit, however also the RMSD value here also keeps stable after around 0.5ns.

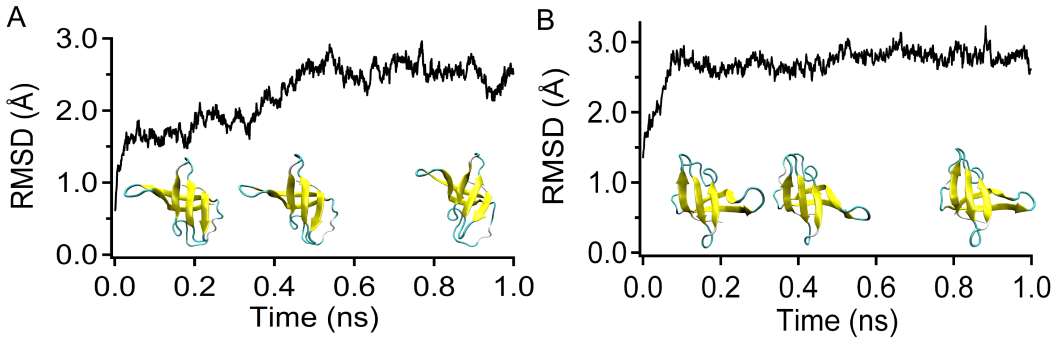

Figure 6.37: Change in the RMSD value during the MD simulation. Corresponding snapshots of the Csp structure at 0.2,0.4 and 0.9ns are shown. (A) Original Csp structure (C_α atom of N-terminus fixed) (B) Turned Csp structure (C_α atom of C-terminus fixed)

All conducted SMD simulations started then with the same coordinate file from the MD simulation.

6.3.2 Constant velocity SMD simulations

Force/extension curves from a total number of five SMD simulations at a constant velocity of 1m/s are shown in figure 6.38. Therein the force applied on the C_α-atom of the C-terminal residue and its coordinate change in x-direction are plotted. The starting x-coordinate (starting extension) of the pulling C_α atom was set to zero in the force/extension curve, when the SMD simulation starts in order to have a better comparison to the experimentally estimated length. All the five SMD simulations shown in different colors in panel A reveal a similar unfolding pattern. In the force/extension curves mainly three peaks are visible, the first at the beginning around 0.5nm (1ns), the second at around 7nm (8ns) and a third peak at around 14nm (14ns). However also slightly differences can be found, e.g the blue trace has not a distinguishable 2nd peak and the black trace not a distinguishable third peak. The purple trace shows additional peaks between the 1st and 2nd peak. At the end the force increases like for the WLC model, when just the elongated and fully unfolded amino acid chain of the Csp is being further stretched. In all SMD simulations conducted at 1m/s the Csp was completely unfolded after around 24ns. Panel B shows one force/extension curve out of the five from the figure in panel A, because it reveals all the three identified peaks. Panel C shows the snapshots of the Csp trajectory taken at the corresponding number indicated in the trace in figure panel B.
From the identified force resistance peaks and the corresponding snapshots, it can be seen that first the β_5 strand gets disrupted from the β_4 (see also the topology of Csp as shown in figure 6.36) from point 1. (0ns,0nm) to point 2. (1.2ns,0.6nm). Both beta strand get

then elongated until the second peak appears, at which the loop gets detached from the remaining inner beta stands β_1-β_3 at point 3 (7.8ns,7.3nm). The last peak arises when the remaining beta strands finally gets disrupted at point 4.(14,5ns,14.1nm). Afterwards the stretched amino acid chain of the Csp remains see point 5. (22.4ns, 21.6nm). However for example the purple trace in panel A indicates additional resistance during the unraveling of β_5 and β_4 from the loop. This indicates a high variety of the simulated unfolding process.

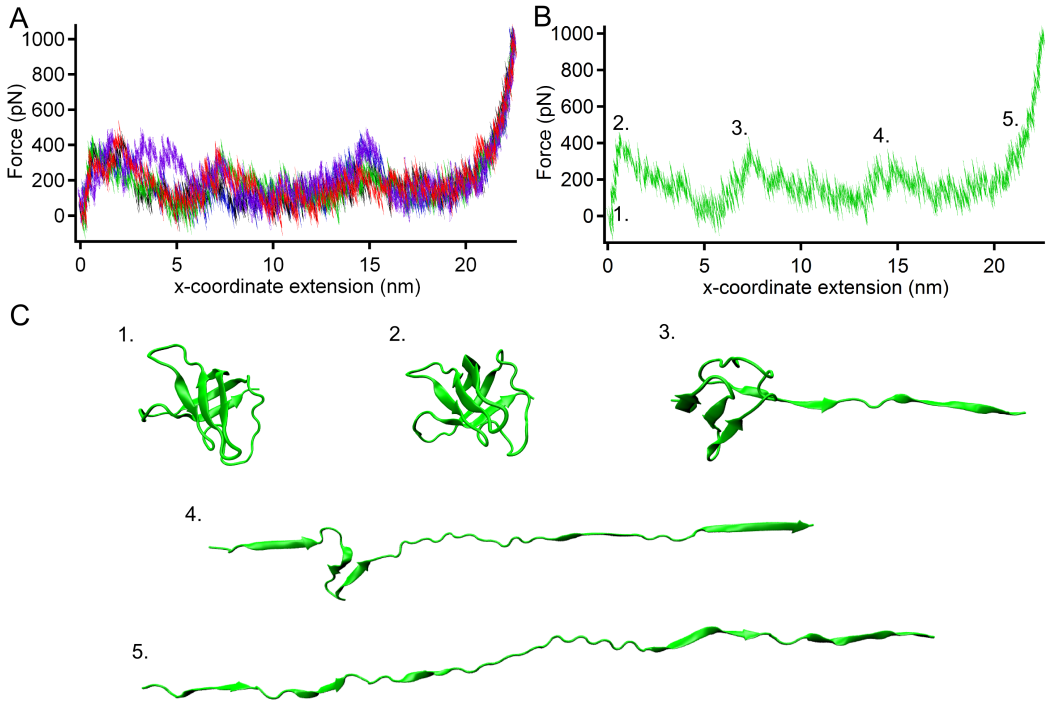

Figure 6.38: A) All five conducted SMD constant velocity traces are shown (B) Single constant velocity SMD trace (C) Snapshots of the SMD trajectory from (B) are indicated with the corresponding numbers from 1. to 5.

All performed SMD simulations show the same unfolding pattern, as described in panel C for the green curve. However as this simulations are conducted at 1m/s, which is still 10^7 times faster than the velocities used in the corresponding AFM experiments (400nm/s), a realistic comparison between the simulations and the experiments are difficult. Also a further decrease in constant velocity to 0.1m/s, which was still bearable computationally to simulate, would not get much closer to the experimental performance. Therefore in the following constant force SMD simulations were carried out, which are closer to the experimental conditions.

6.3.3 Constant force SMD simulations

Figure 6.39 shows all the Csp separation length vs time trajectories from the performed constant force SMD simulations. Therein the separation length is the length measured between the C_α-atom of the N-terminal residue and the C_α-atom of the C-terminal residue.

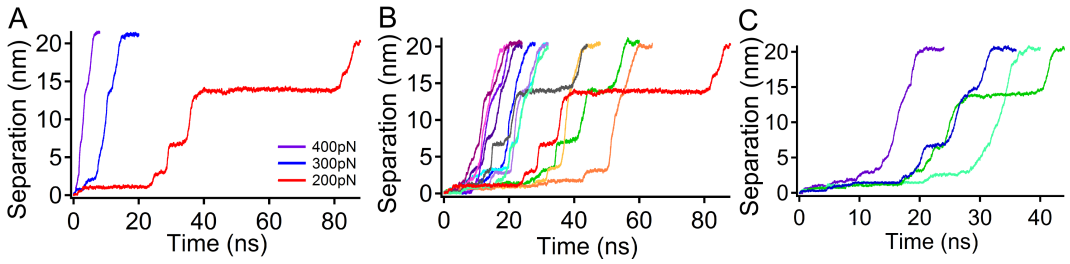

Figure 6.39: (A) Constant force SMD simualtions at 200, 300 and 400pN (B) Thirteen constant force SMD simulations at 200pN (C) Four constant force simulations at 200pN, but in comparison to B with the changed direction of pulling.

In panel A three trajectories from SMS simulations at different constant forces 200, 300 and 400pN are compared. Although the trace at 200pN unfolds after 88ns, it has been found that this force is an ideal force to conduct SMD simulations on Csp, as it unfolds in a reasonable calculation time, which can be performed by the used computational resources. Additionally the probability of observing intermediates with a long dwell time (up to 20-40ns) is increased at a constant force of 200pN (red trace). Panel B shows the variety of thirteen separation length vs time traces, all performed at a constant force of 200pN. Although all simulations started from the same coordinate structure file, the Csp unfolding trajectory can appear like a single step, whereas also multiple steps can be observed. Panel C shows the same variety like in panel B of four separation length vs time traces, all carried out at a constant force of 200pN, but with the opposite pulling direction as in panel B and using the equilibration MD coordinate file as shown in figure 6.37 B.

Figure 6.40 illustrates the details of the red trajectory from figure 6.39 panel A and B. Panel A shows the single separation vs time trajectory, which indicates six clear distinguishable intermediate steps. The overall unfolding time is 88ns. Panel B shows the corresponding secondary structure content plotted using the timeline plugin of the VMD program. Together with the snapshots shown in panel C, the whole unfolding trajectory can be followed. The unfolding pathway of this trajectory is very similar to the one observed during the constant velocity simulations (see figure 6.38C). The unfolding starts by opening the two beta strands β_5 and β_4, which happens between 1-3ns. The structure maintains at this position until around 20ns when the β_5 strand gets further elongated

until around 24ns where it stays until around 27ns. Here the β_4 strand gets disrupted from the loop and totally straightened, which happens between 27 and 30ns. At this point the Csp keeps 4ns until also the loop gets detached from the remaining three beta strands (β_1-β_3). At this intermediate, as it can be seen in the snapshot done at 40ns, the Csp keeps around 40ns. Then at around 84ns the β_1 strand breaks and the whole amino acid chain gets elongated.

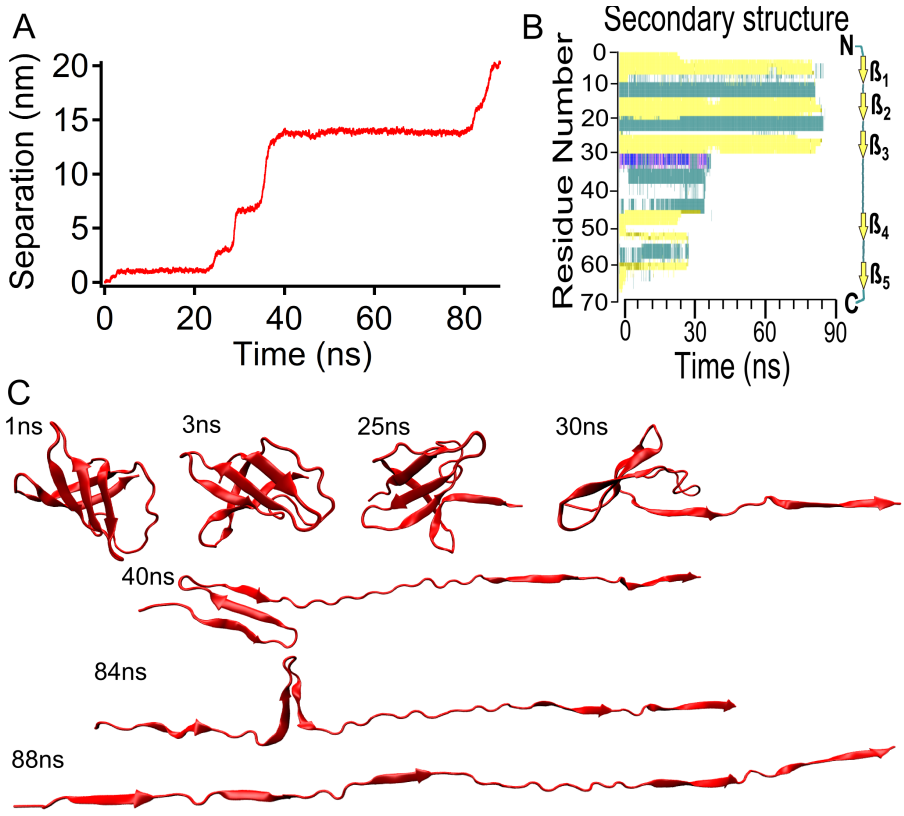

Figure 6.40: (A) Constant force SMD simualtions at 200pN (B) Secondary structure content vs time (C) Snapshots from the trajectory of the Csp, taken at the indicated time.

For comparison figure 6.41 illustrates the details of the blue trajectory from figure 6.39 B. Here the single separation vs time trajectory (panel A) does not indicate any clear distinguishable intermediate steps. Stable intermediate states are not detectable and therefore the Csp unfolds through one single step. The overall unfolding time is 28ns. Panel B and C again facilitate to follow the whole unfolding trajectory. Although the unfolding starts like in the red trace in figure 6.40 by opening the two beta strands β_5 and β_4, which happens between 4 and 8ns, the unfolding pathway of this trajectory is very different. Here first the β_1 strand gets detached from the beta strands β_2 and β_3, which happens between 16 and 19ns. Then the beta strands β_5 and β_5 are elongated further for another ns. Afterwards the loop and the rest of the amino acid chain stretched one after the other,

until the Csp is fully unfolded around 25ns.

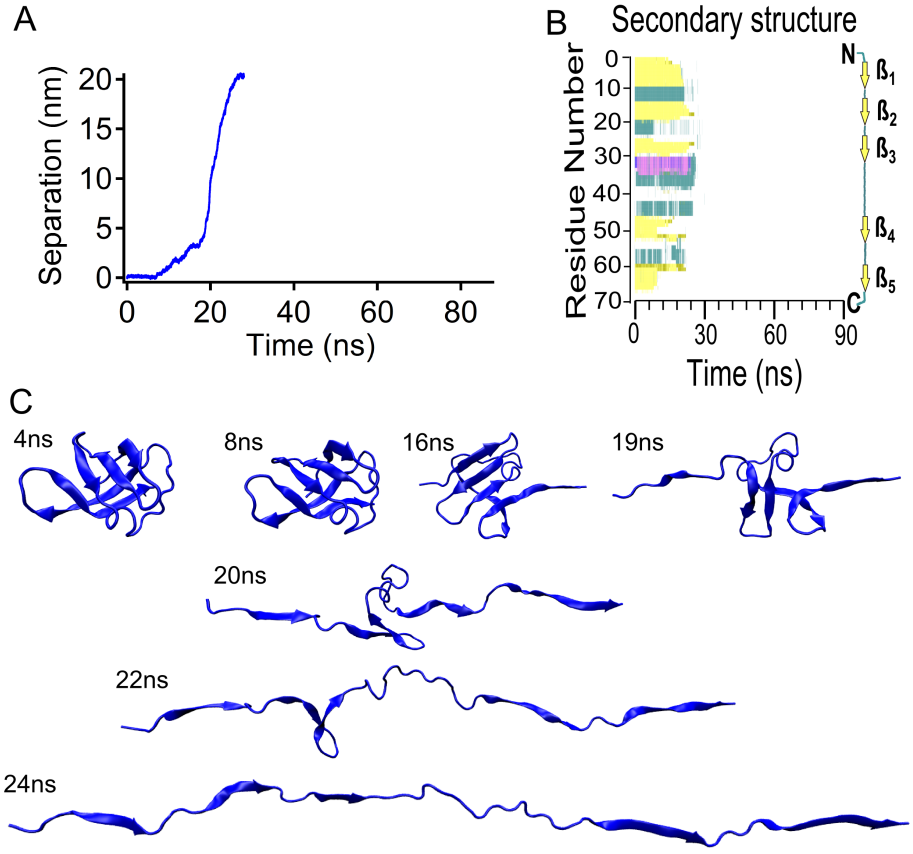

Figure 6.41: (A) Constant force SMD simualtions at 200pN (B) Secondary structure content (C) Snapshots from the trajectory of the Csp, taken at the indicated time.

The third example in figure 6.42 illustrates the details of the green trajectory from figure 6.39 C in order to have an example of a simulation with changed pulling direction. The single separation vs time trajectory does indicate two clear distinguishable intermediate steps, the first around 26ns and the second around 40ns (panel A). Herein one stable intermediate state is detected for around 13ns. The overall unfolding time around 44ns. Panel B and C show the corresponding content of secondary structure and the snapshots from the trajectory. Although during this simulation the force is applied on the opposite side compared to the simulations before, the overall unfolding behavior is like the one discussed for the red trace in figure 6.40. Also here the unfolding starts by opening the two beta strands β_5 and β_4, which happens between 4 and 16ns. Between 20 and 24ns the β_5 and β_4 strands gets fully elongated. Until 28ns also the loop gets straightened. The Csp now stays at this intermediate for around 13ns until the beta strands β_1 gets detached from the inner beta strands β_2 and β_3. Afterwards the amino acid chain gets fully unfolded around until around 44ns.

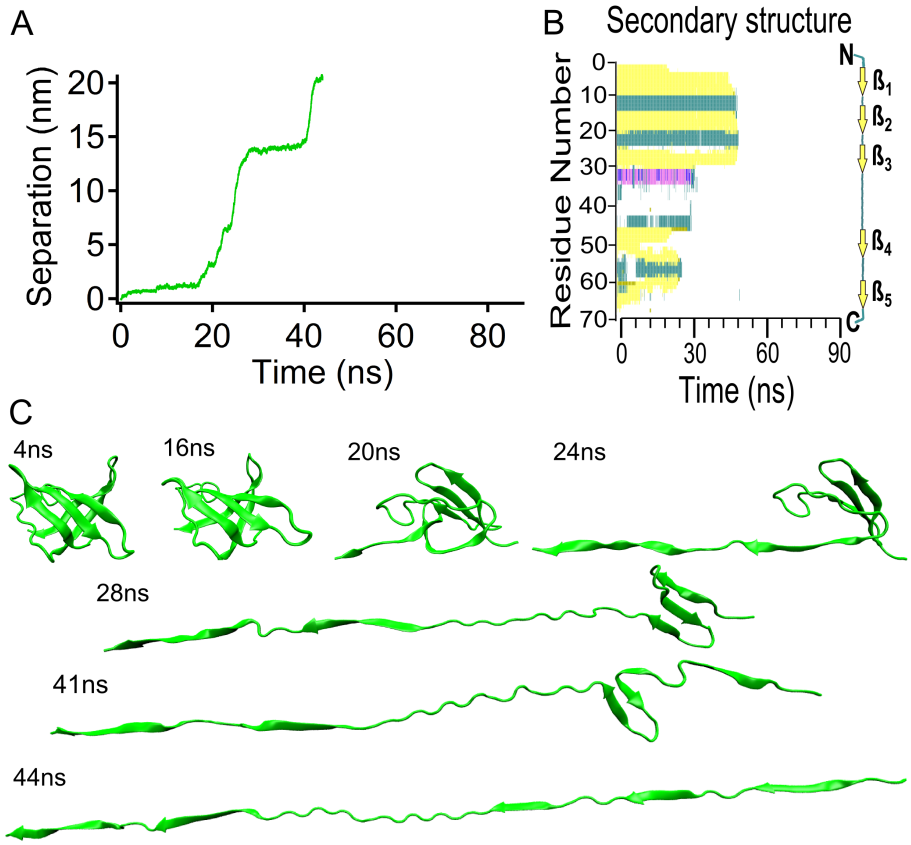

Figure 6.42: (A) Constant force SMD simualtions at 200pN (B) Secondary structure content (C) Snapshots from the trajectory of the Csp, taken at the indicated time.

Figure 6.43 compares the three previously discussed unfolding trajectories (panel A). After around 25ns the green and the red trajectory follow the same trend, although in the red trace the Csp unfolds completely at a much later time.

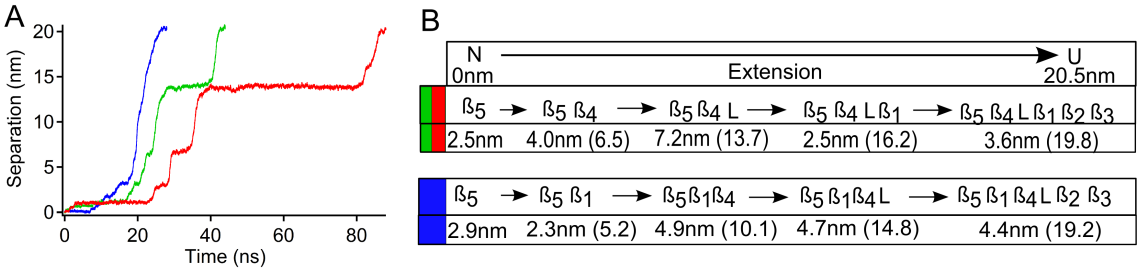

Figure 6.43: A) The three discussed constant force SMD simulations at 200pN (B) Summary of the detected unfolding pathways of the three simulations, which indicate the order of the unfolding of secondary structure elements and the corresponding change in length

The table in panel B, shows the change in length during the detected three unfolding pathways of the three simulations. As the unfolding pathway of the red and green trace were the same, here the average values have been used. The table indicates not only the single step lengths, furthermore also the absolute lengths are shown in brackets.

Summary and Discussion

The conducted SMD simulations describe the unfolding pathway of the Csp at atomic resolution. The idea was to determine whether the simulations could reproduce the experimental observations and therefore provide further detail about the unfolding process of the Csp. Therefore also the effect of applying a constant force from the opposite direction was investigated as during the experiment it is suggested that the force is applied on one or the other end of the protein of investigation by a probability of 50%.

The main result from the performed SMD simulations is that they indeed can reproduce the experimental observations at least in relative terms as the unfolding process during the simulations takes place many magnitudes faster than in the experiment. It has been found, that the unfolding pathways can vary, especially under a constant force from a single step to a multiple step unfolding behavior of the Csp and also in the observed order of releasing secondary structure elements due to unfolding. Furthermore the observed step lengths varied between 2 and 20nm, which is the total unfolding length. However changing the sites of applying a mechanical force did not result to influence the overall unfolding behavior as the same unfolding pathways were observed. Figure 6.44 lists all the detected unfolding pathways and the corresponding change in lengths from the seventeen conducted 200pN constant force SMD simulations.

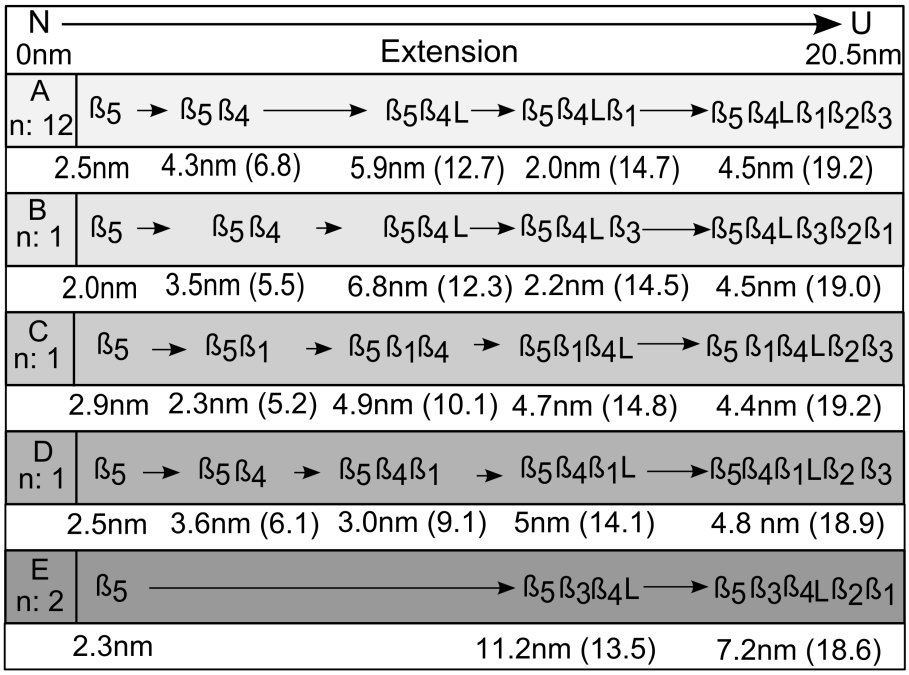

Figure 6.44: Summary of all detected Csp unfolding pathways during constant force SMD simulations at 200pN

Five different unfolding pathways could be estimated from A to E, which all start from the same point wherein the β_5-strand gets disrupted from β_4 strand. The green and red curve from figure 6.43 corresponds to the unfolding pathway A and the blue trace from figure 6.43 corresponds to the pathway C. From the seventeen trajectories twelve follow the pathway A, which is also been identified in all the five conducted 1m/s constant velocity SMD simulations (see figure 6.38). Another two 200pN constant force simulations have been identified to follow the unfolding pathway E, in which the β_5 and the inner β_3-strand get detached simultaneously.
Compared to the found clusters C1 to C8 (see figure 6.29 and to the experimentally estimated distribution of unfolding step lengths during constant force measurements (see figure 6.28), the lengths of the unfolding of the structural elements within the Csp from figure 6.44 give an additional picture of the mechanical unfolding process of Csp. Together with the variability in unfolding pathways this could explain first the high variability of experimentally measured unfolding lengths and second the missing of any consecutive and force dependent unfolding order of secondary structural elements.

6.4 Conclusions

Single molecule force spectroscopy measurements with the force-clamp AFM unveiled that the mechanical unfolding mechanism of the two-state folder Csp is highly heterogeneous. This heterogeneity results from the variety of detected individual unfolding behaviors that ranges from single step unfolding to unfolding events that consists of multiple intermediates, some of which remain stable for many seconds during the unfolding experiments. The most remarkable finding was that this variety of Csp unfolding behavior is controllable or can be modulated with the amount of pulling force (constant force), wherein the maximum of heterogeneity in unfolding pathways of the Csp is reached at forces between 40-60pN. Furthermore the observed heterogeneity is of such complex nature that it cannot be explained with a simple sequential unfolding pathway in which the structural elements of Csp unfold one by one following a defined order of step lengths and unfolding forces. The performed cluster analysis of the AFM experimental data and SMD simulations affirm this finding that the unfolding process of the Csp occurs via multiple independent pathways, which can be resolved on the single molecular level. It seems that just by the application of a well defined reaction coordinate like a mechanical force, which tilts the global free folding energy surface to the unfolding state, can reveal these otherwise hidden multiple pathways.

This direct observation of such kinetic heterogeneity during mechanical unfolding experiments is quite a remarkable finding because reported protein folding experiments (in bulk and on the single molecule level) have almost in every case produced a simple two-state observation.

The existing experimental reports of complex unfolding/folding behaviors are always associated to conformational processes that are different to two-state processes. E.g. single molecule force spectroscopy experiments on the three-state folding T4 lysozyme [133], the molten-globule-like unfolding of a membrane-associated protein [170], and the mechanical expansion of unfolded polyubiquitin chains that were previously collapsed by a force-quench pulse [151], have all reported a very broad distributions of unfolding lengths. Similarly, the collectively simple bulk thermal unfolding of one-state downhill folding BBL domain [83] and the ultrafast folding protein gpW [78, 35] have been shown to become very complex when investigated at atomic resolution.

However the Csp folding rate is far away from the ultrafast folding regime that is predicted to result in such marginally cooperative unfolding processes [58]. Instead when studying the Csp with single molecule resolution using both single molecule fluorescence (FRET) [73, 67] and force extension AFM [128] Csp has always robustly shown a simple two-state-like behavior.

It is only with the application of moderate, precisely controlled, pulling forces in single-molecule mechanical unfolding experiments using the force clamp AFM that the remarkably heterogeneous features of Csp unfolding become apparent and thus directly comparable to atomistic computer simulations. A recently published study describes a very similar observation on the two-state folding protein src SH3 [203]. Therein an approach combining mechanical and chemical unfolding with mutational analysis was applied to obtain evidence that the src SH3 unfolds via multiple unfolding pathways that are differentially affected by the perturbation. This suggests that un(folding) through multiple pathways may be in fact a more general observation for two-state folding proteins.

Thus, force-clamp AFM in combination with SMD emerged a perfect combination for probing at high resolution the topographic features of the free energy landscapes of the two-state folding protein Csp and how it respond to force perturbation. For the first time the observation of multiple pathways became so clearly by using a single molecule technique like the force clamp AFM, which is confirmed by SMD simulations.

Furthermore the performed stochastic kinetic simulations reveal a good agreement with the experimental results both force clamp and constant force by just using the values for

the kinetic un(folding) rates at zero force and the estimated distance to the unfolded state by applying the Bell model. Therefore this simple model might be very useful to interpret experimentally obtained single molecule force spectroscopy data. On the other hand this simple simulation can be used to plan mechanical unfolding experiments, especially to have an idea what forces, unfolding lengths or dwell times are expected depending on the protein of investigation if the kinetic values are already known.

Now at this point all the experimental and computational techniques were applied to the ultra fast folding protein gpW to investigate the effect of a higher un(folding) kinetic rate.

Chapter 7

Mechanical unfolding of the gpW protein

This chapter contains the following sections, which will provide the overall results describing the mechanical behavior of gpW. The first sections cover all conducted SMFS experiments on gpW, including constant velocity, force ramp and constant force AFM measurements. Then the results of the SMD simulations on gpW will be summarized.

7.1 SMFS measurements on the gpW protein

As in the case of SMFS measurements on BBL and Csp, just traces were collected, in which at least four titin I27 unfolding events could be detected. Furthermore traces were discarded applying the same criteria as discussed in the two chapters before (see figure 5.1) for constant velocity, figure 6.5 for force ramp and figure 6.5 for constant force measurements. In the case of the gpW protein the unfolding length before the first I27 domain unfolds is very similar to the one for the Csp as it has almost the same amount of amino acids. The expected unfolding length of the gpW is around 20,4-26nm (51-65 amino acids), depending on if the amino acids from the tail at the N and C-terminus are included or not. With the estimated length of the six folded I27 domains (26.4nm), the first I27 domain should unfold after a length between 20,4-26nm and 46.8-53.4nm.

7.1.1 Constant velocity measurements

As in the constant velocity measurements performed before at the beginning of the trace the cantilever was first moved against the surface with a force of $|\tilde{F}| = 1$nN, before the cantilever retracts from the surface. Again in each force/extension trajectory depicted the

red trace is the approach and the blue trace is the retraction of the cantilever from the surface. The insets in each trajectory figure show the beginning of the trace at a higher magnification to illustrate the unfolding behavior of gpW, as the gpW protein most likely has a lower unfolding force than the one from the titin I27 domain.

The unfolding force at each peak was directly measured and the difference in contour length was estimated with fitting each peak to the WLC model. The measured values of ΔL of gpW and titin I27 will be indicated in the figures.

Constant velocity measurements at 400nm/s

Figure 7.1 shows four typical force extension traces measured at a constant velocity of 400nm/s with the MultiMode AFM.

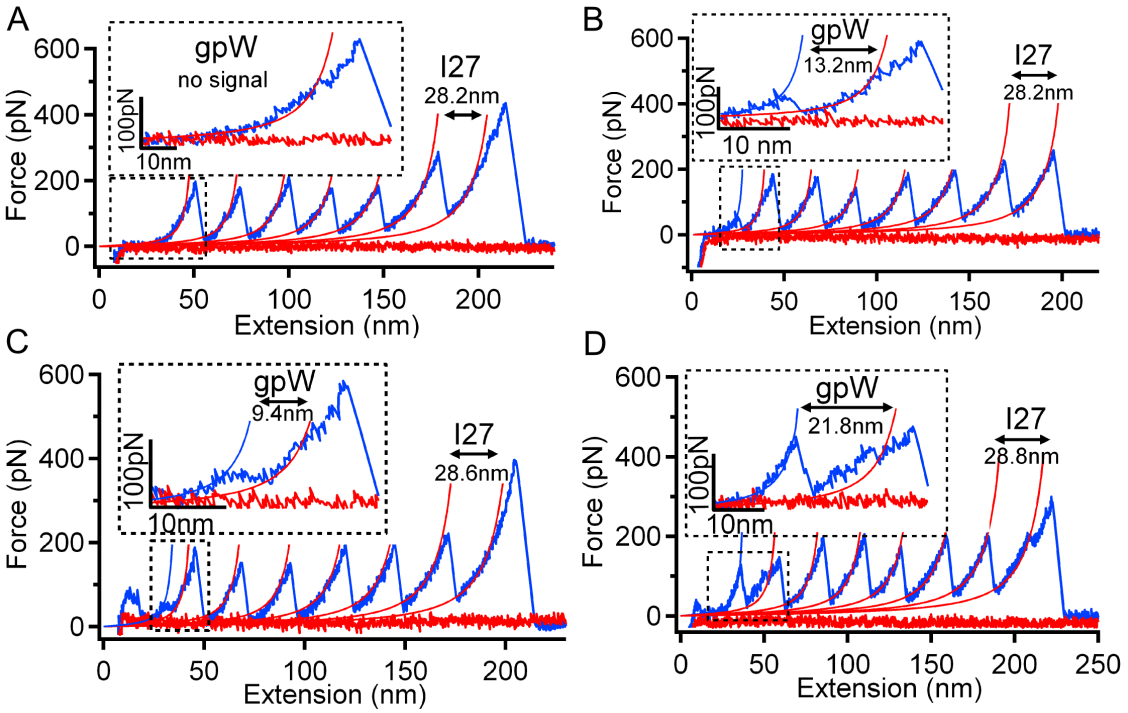

Figure 7.1: Force extension traces for gpW at a constant velocity of 400nm/s (A) No gpW unfolding signal is detected (B) gpW unfolds via a single peak (C) gpW unfolds via a hump (D) A very heterogeneous unfolding signal of gpW

All four depicted traces show the complete $I27_3$-gpW-$I27_3$ unfolding, because each one depicts six titin I27 unfolding peaks. The unfolding pattern of gpW was found to be very heterogeneous within the selected traces (n=138) and shows exactly the same characteristics, which has been observed on the BBL domain (see figure 5.2). Therefore the observed heterogeneity of gpW was clustered into the same three groups, which were observed for the BBL domain. In the case of gpW around 34% of all traces showed no gpW unfolding

signal (see figure 7.1 A). 44% of all traces showed a gpW unfolding behavior similar to a unfolding peak (see figure 7.1 B). In the rest of 24% traces the gpW appeared to unfold through a hump (see figure 7.1 C). Some traces of gpW revealed a high degree of heterogeneity, in which the trajectory was difficult to distinguish between a single peak or a hump (see figure 7.1 D). Therefore and especially in these cases just the beginning of this signal was fitted to the WLC model to approximate the total unfolding length of the gpW protein as it has been also applied on the BBL before. Also the unfolding force was just measured at the point of fitting.

The measured distributions of unfolding force and difference in contour length of gpW and for the titin I27 domain are given in figure 7.2.

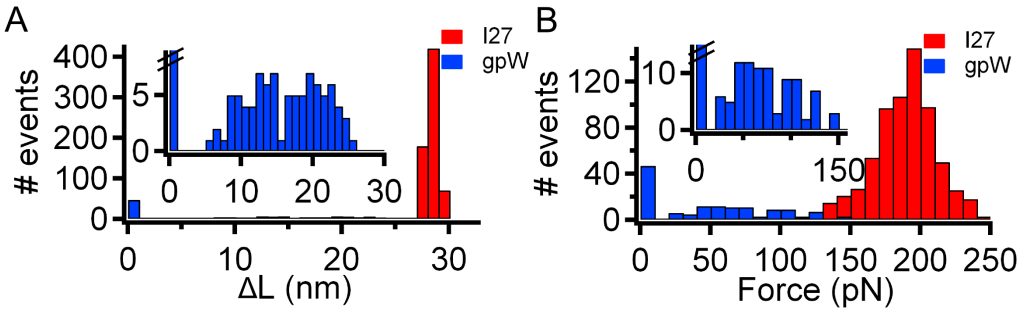

Figure 7.2: Histograms of the estimated distributions of the estimated difference in contour length (bin size 1nm) (A) and the measured unfolding force (bin size 10pN) (B) of the gpW protein and the titin I27 domain at a constant velocity of 400nm/s

Averaging the distributions for the gpW results in a mean unfolding force of (49.1 ± 43.7)pN and in a mean difference in contour length of (10.5 ± 8.7)nm. The corresponding results from the average of the titin I27 domain are (188.9 ± 23.4)pN and (28.3 ± 0.5)nm, which coincide with the results for the I27 from the corresponding $I27_3$-Csp-$I27_3$ and $I27_3$-BBL-$I27_3$ measurements.

Summary and Discussion

Force extension measurements on the $I27_3$-gpW-$I27_3$ construct with a velocity of 400nm/s revealed a very heterogeneous unfolding behavior, being very similar to the results obtained when measuring the $I27_3$-BBL-$I27_3$ sample. However some differences were observed. A slightly higher percentage of traces showing no gpW unfolding signal were detected ($\approx$ 34%) compared to corresponding measurements of the single BBL domain ($\approx$ 25%). The measured ΔL distribution of gpW was slightly shifted to higher values, which was expected as gpW contains more amino acids than the BBL domain. The average of the measured unfolding force of gpW is lower than estimated for the BBL, because more

traces with no gpW signal were counted compared to the BBL measurements. This shift the average force to lower values. However, when a signal has been detected its estimated distribution in unfolding force was very similar to the corresponding distribution of the BBL.

In sum the measured difference in contour length for gpW revealed a broad distribution from 5 to 25nm. Here it has to be pointed out that a difference in contour length below 5nm is difficult to estimate and to clearly separate from the WLC fitting of the first I27 unfolding peak. Therefore, when adding the traces with no measured gpW signal (47 traces), the center of the measured length distribution would change to the lower length values from 0-5nm (see panel A in figure 7.2).

The corresponding unfolding force distribution covers the whole range from the AFM detection limit 15-20pN to 150pN. Therein, if the traces with no detectable gpW signal are added (46 traces), the center of the distribution lies between 0-20pN (see panel B in figure 7.2).

However as the amount of traces in which no gpW unfolding signal can be measured is high, it is suggested that gpW cannot be properly resolved when using the MultiMode AFM. Furthermore due to the heterogeneous appearance of the gpW unfolding pattern it becomes difficult to distinguish clean unfolding traces from unfolding traces, which could contain noise or fluctuations caused by interactions between the AFM cantilever, the protein sample and/or the gold substrate. This could cause also the slightly too broad estimated difference in contour length distribution or a unfolding force distribution with too high values. Here for example the issue of distinguishing between noise from no noise containing traces was much easier to resolve in the case of the Csp, as it unfolds every time through a clear unfolding peak.

In order to get more insights into the mechanical unfolding behavior of the gpW sample and especially to explore the low unfolding length and force regime, controlled force ramp and constant force measurements were conducted using the force clamp AFM.

7.1.2 Force ramp measurements

The force ramp measurements were conducted in the same conditions as for the Csp. At the beginning of the trace the cantilever was pushed first against the surface with $|\tilde{F}| = 1$nN. This value was then increased within 0.1 s to -20pN, from which the force was then linearly increased with the corresponding measuring rate up to 300pN in order to detect the titin I27 unfolding steps in the end of the trace. Here in some case the ramp was

just set until 150pN, as the unfolding behavior of the I27 protein was already confirmed and will be cited during the corresponding section. As before, in each trajectory shown the red trace is the length vs time and the black the corresponding force vs time trace. The insets show the section of trajectory which belong to the gpW at a higher magnification to illustrate its unfolding behavior in particular indicating the corresponding length and time scale.

During the force ramp measurements it was observed that in contrast to the Csp, the gpW is not unfolding via a sharp single or well defined multiple steps. Moreover gpW was detected to unfold via a broad sigmoidal shape curve. Hence its total unfolding length was roughly estimated by measuring the length difference between the beginning and the end of its sigmoidal shaped unfolding signal. The corresponding force value (mid force) was measured in the middle of the sigmoidal shape signal to serve as a first approximation. This value is not to be confused with the force $F_{1/2}$ value, which describes the force at which the un(folding) rates under a mechanical force are equal. Thus, the unfolding mid force and the corresponding length of the gpW and the titin I27 domain was measured directly using the cursors in each trajectory.

Furthermore all force-clamp measurements on gpW were conducted with the BioLever cantilever (cantilever B, spring constant k $\sim$ 3-6pN/nm) as it lower spring constant enables a sensitive measurements in the low force regime.

Force ramp measurements at a force rate of 20pN/s

As it has been observed during the constant velocity measurements, the force ramp experiments of gpW revealed also a high heterogeneity in unfolding patterns. Four examples of force ramp traces at a constant force rate of 20pN/s are illustrated in figure 7.3. All four depicted traces show the complete unfolding of the polyprotein with the gpW signal and six titin I27 unfolding steps. From all measured traces (n=78) the gpW unfolding pattern was divided into three groups.

Within the first classified group (n=8 traces) the gpW unfolds smoothly via a sigmoidal shape unfolding curve (panel A). Therein fluctuations in length during the sigmoidal shape transition were detected which can vary between 3-5nm. The majority of traces have been found to be in the second group (n=58), in which the gpW unfolds via several steps within the sigmoidal shape unfolding signal (panel B). Furthermore traces were detected in which the gpW unfolds via a single step (n=12; panel C), which was classified to be the third group of detected traces. However additionally traces are observed, in which the cantilever

moves slightly back and forth (panel D) within a range of 4-5nm inside the gpW sigmoidal shaped unfolding pattern.

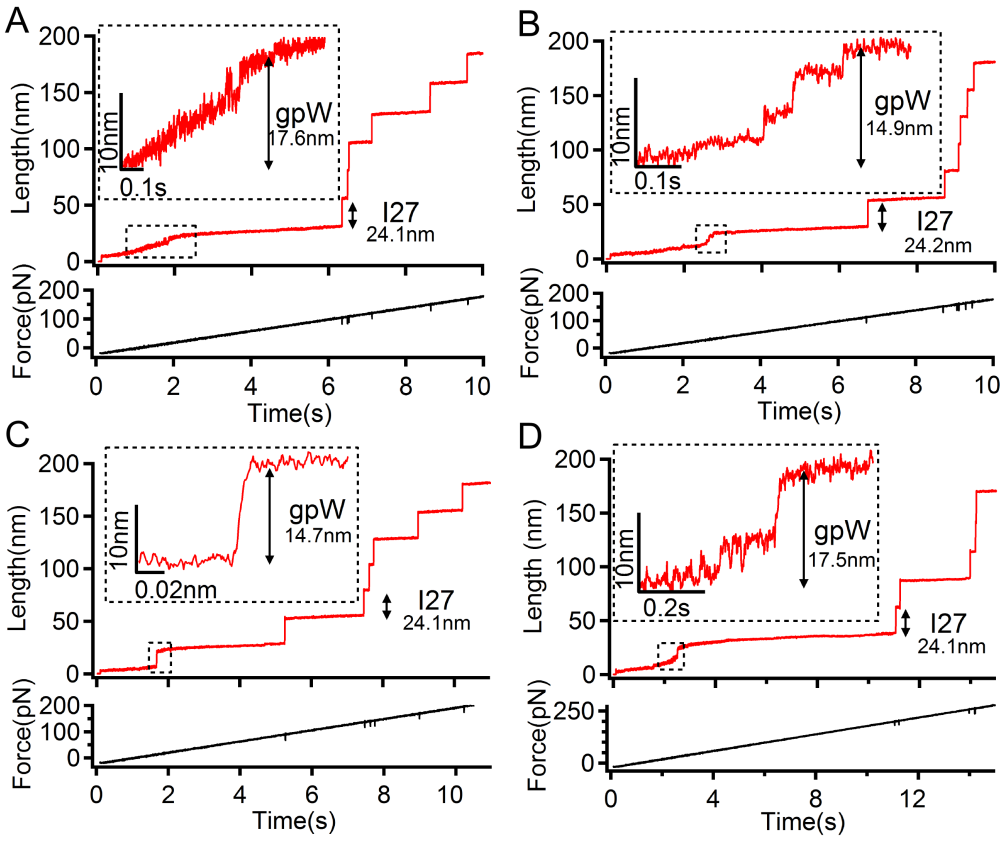

Figure 7.3: Force ramp traces for gpW at a constant force rate of 20pN/s (A) gpW unfolds through a sigmoidal shape curve (B) gpW unfolds via a several steps but still inside a sigmoidal shaped curve (C) gpW unfolds via a single step (D) With using the BioLever cantilever larger fluctuations (>5nm) in length are visible

Figure 7.4 gives an overview of the measured distributions of unfolding force and length of the gpW and for the titin I27 domain. The average of the distributions for the gpW results in a mean unfolding length of (17.5 ± 2.5)nm and in a mean unfolding mid force of (14.7 ± 9.5)pN. The average of the distributions of the titin I27 domain are (24.4 ± 0.6)nm and (140.8 ± 26.5)pN.

The effect of increased force resolution when using the Biolever cantilever can be seen in the unfolding force value obtained for the I27 domain, which is around 20pN lower than the corresponding value obtained with the MLCT cantilever for the Csp 6.4. Very interestingly during these measurements traces were detected like in panel D in figure 7.3, in which the large fluctuations (<5nm) were detected.

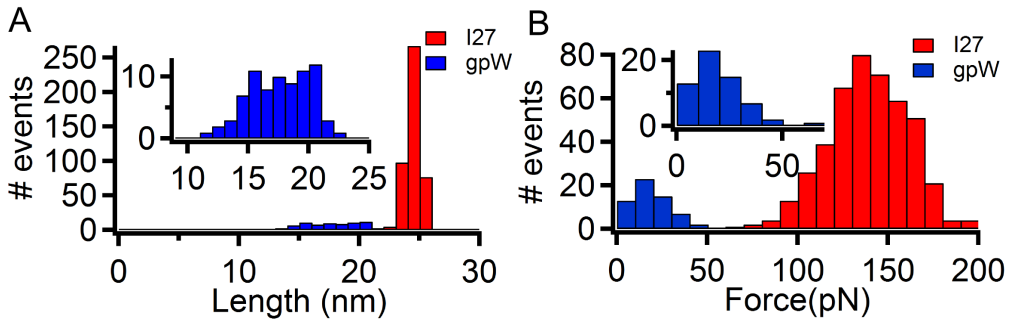

Figure 7.4: Histograms of the distributions of the estimated unfolding length (bin size 1nm) (A) and the measured unfolding force (bin size 10pN) (B) of the gpW and the titin I27 domain at a constant force ramp of 20pN/s.

Force ramp measurements at a force rate of 40pNs

All in all no huge difference could be detected in the unfolding behavior of the gpW between 40 and 20pN/s. Figure 7.5 shows one example for each of the three groups (sigmoidal shape curve, sigmoidal shape curve with steps and a single step) of force ramp traces detected at a force ramp of 40pN/s.

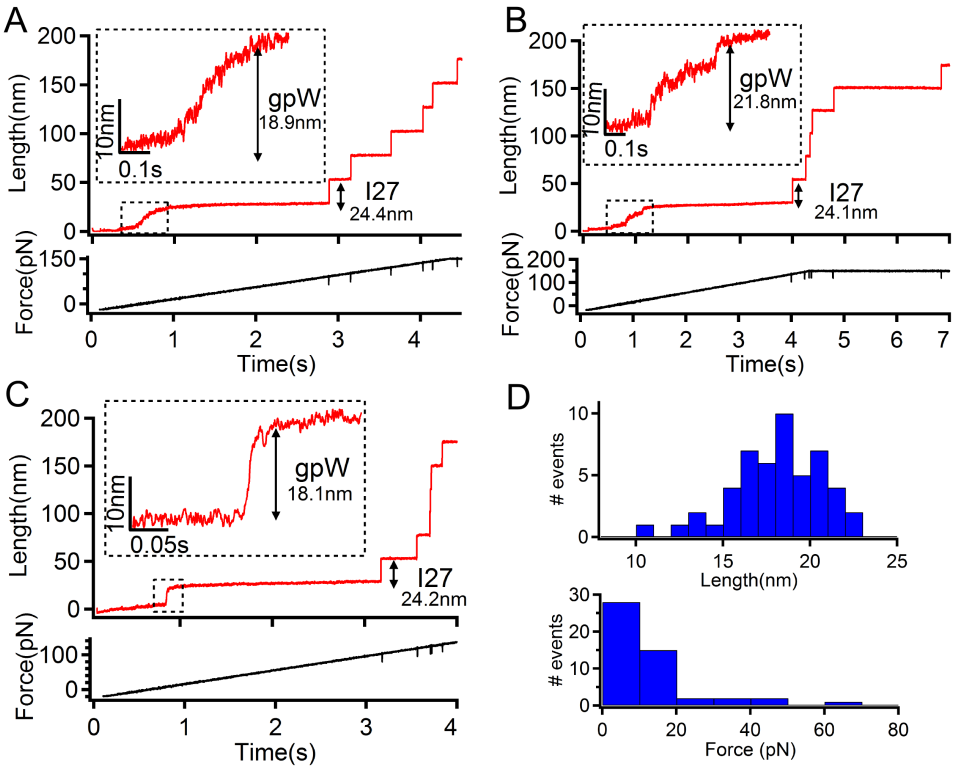

Figure 7.5: Force ramp traces for gpW at a constant force rate of 40pN/s (A) gpW unfolds through a sigmoidal shape curve (B) gpW unfolds via a several steps but still inside a sigmoidal shaped curve (C) gpW unfolds via a single step (D) Histograms of the distributions of the estimated unfolding length (bin size 1nm) and the measured unfolding force (bin size 10pN) of the gpW.

From a total number of traces (n=50) n=19 were counted to unfold within a sigmoidal shape like in panel A. The majority of traces show a similar shape like in panel B (n=26) and a few traces (n=5) were observed, in which gpW unfolds via a single step. The corresponding distributions of gpW (panel D) give an average value of the unfolding force of (12.8 ± 12.6)pN and an unfolding length of (18.0 ± 2.6)nm, which are in the same range like measured at a force ramp speed of 20pNs.

Force ramp measurements at a force rate of 60pNs

Also at a force rate of 60pN/s, a huge difference could not be detected in the unfolding behavior of the gpW compared to the measurements done before at 40 and 20pN/s. One typical occurring trace for each of the classified three groups (sigmoidal shape step, sigmoidal shape step with steps and a single step) at a rate of 60pN/s are depicted in figure 7.6.

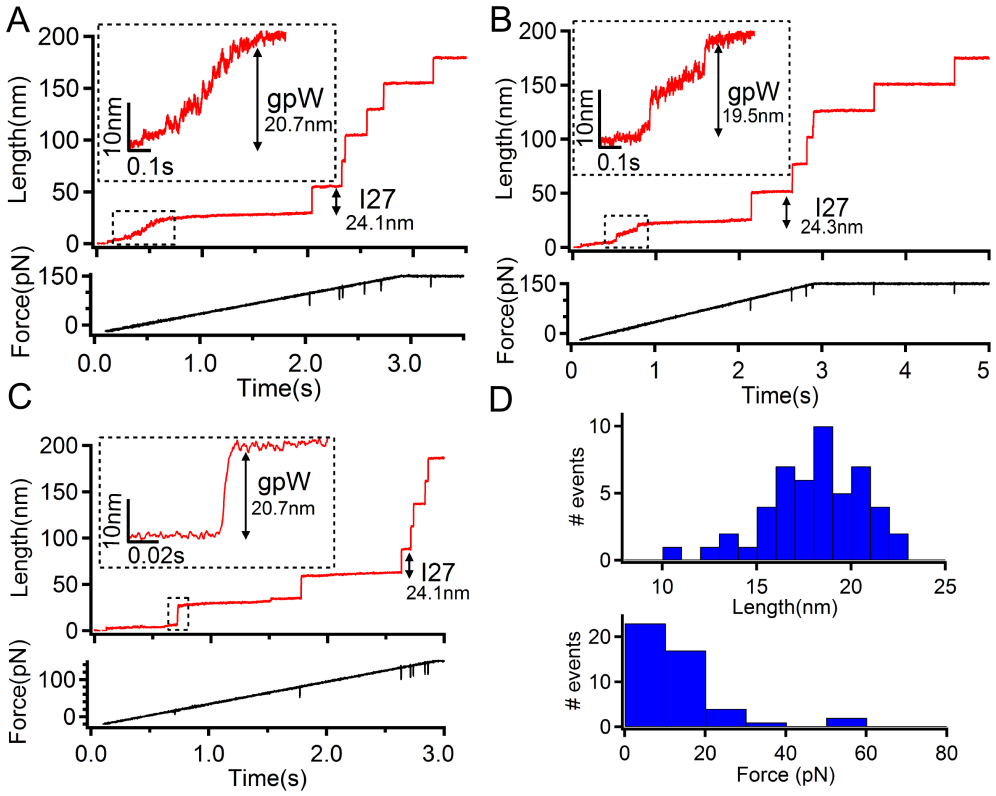

Figure 7.6: Force ramp traces for gpW at a constant force rate of 60pN/s (A) gpW unfolds through a sigmoidal shape curve (B) gpW unfolds via a several steps but still inside a sigmoidal shaped curve (C) gpW unfolds via a single step (D) Histograms of the distributions of the estimated unfolding length (bin size 1nm) and the measured unfolding force (bin size 10pN) of the gpW.

From a total number of traces (n=47) n=17 were counted to unfold within a sigmoidal shape (see figure 7.6 A). The majority of traces show a similar step shape like in panel

B (n=25) and a few traces (n=5) were observed, in which gpW unfolds via a single step. The corresponding distributions of gpW (panel D) give an average value of the unfolding force of (12.7 ± 11.5)pN and an unfolding length of (18.3 ± 2.5)nm, which are in the same range like measured at a force ramp speed of 40 and 20pNs.

Force ramp measurements at a force rate of 80pNs

No further change in unfolding behavior of the gpW protein could be observed at a force ramp sepped of 80pN/s. Figure 7.7 shows one example for each of the cluster (sigmoidal shape step, sigmoidal shape step with steps and a single step) of force ramp traces detected at a force ramp of 80pN/s.

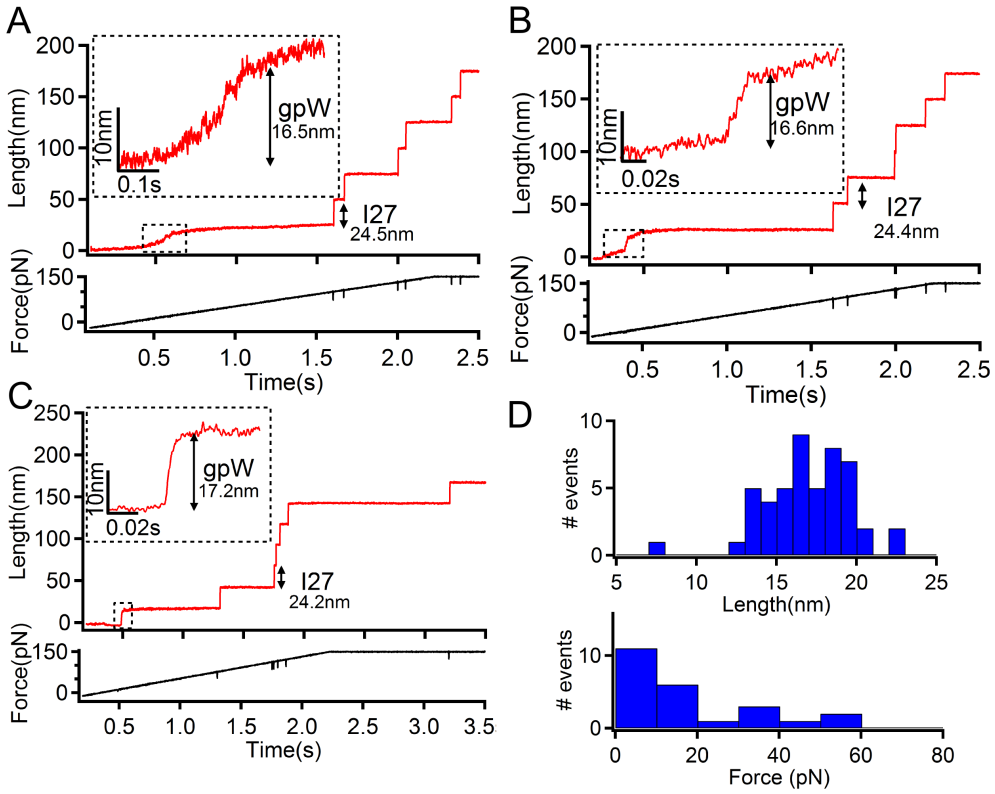

Figure 7.7: Force ramp traces for gpW at a constant force rate of 80pN/s (A) gpW unfolds through a sigmoidal shape curve (B) gpW unfolds via a several steps but still inside a sigmoidal shaped curve (C) gpW unfolds via a single step (D) Histograms of the distributions of the estimated unfolding length (bin size 1nm) and the measured unfolding force (bin size 10pN) of the gpW.

From a total number of traces (n=24), the majority of traces n=14 were counted to unfold within a sigmoidal shape like in panel A. N=6 traces were detected to show a similar step shape like in panel B and a few traces (n=4) were observed, in which gpW unfolds via a single step. The corresponding distributions of gpW (panel D) give an average value of the unfolding force of (18.0 ± 16.2)pN and an unfolding length of (16.7 ± 2.6)nm. Here

the detected unfolding force is a bit higher than at 20,40 and 60pN/s, however the number of total traces is just n=24 and therefore the corresponding unfolding force distribution cannot be directly compared.

At this stage it was now thought to decrease the force ramp speed, because at a higher force ramp speed than 80pN/s the time resolution of the instrument is easily be reached for a protein like gpW, which unfolds at forces between 1-20pN.

Force ramp measurements at a force rate of 1pN/s

The force ramp measurements with the BioLever B enabled stable measurements for a long time at a rate near equilibrium at 1pN/s. At that force ramp speed the titin domains start to unfold after 100-120s. That low force ramp speed made these experiments highly sophisticated but when using the force-ramp AFM stable measurements could be successfully performed. Figure 7.8 shows four examples of force ramp traces at a constant force rate of 1pN/s.

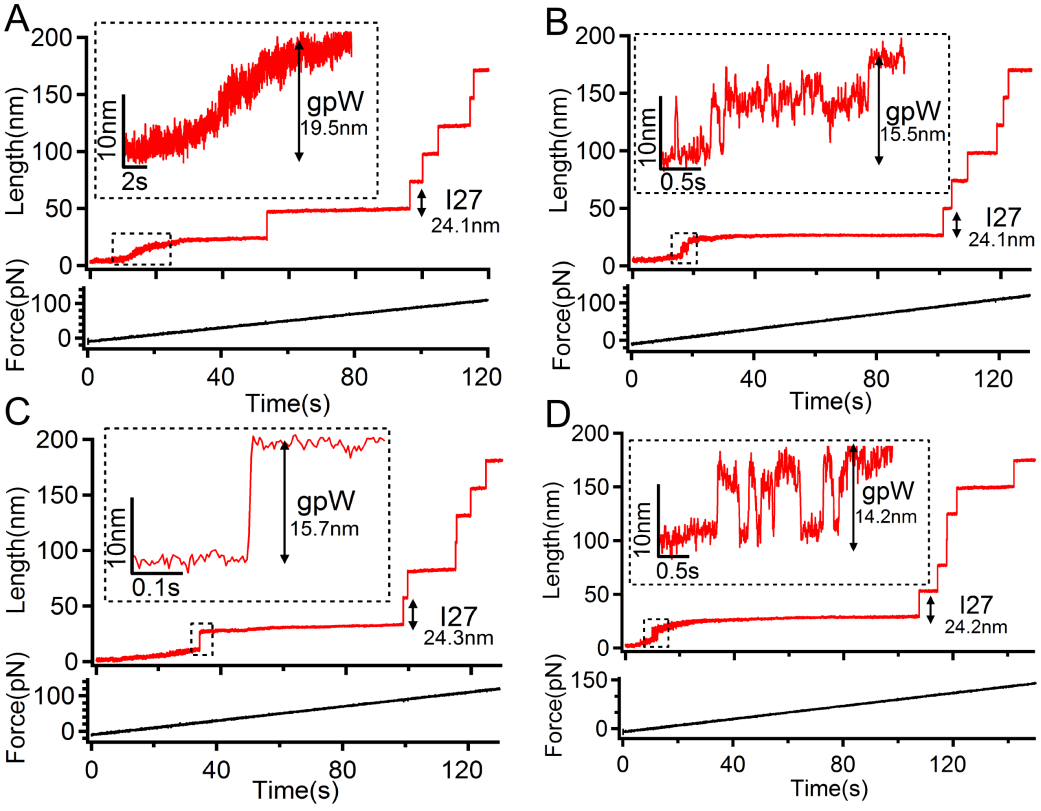

Figure 7.8: Typical force ramp traces for gpW at a constant force rate of 1pN/s near equilibrium. (A) The typical sigmoidal shape unfolding of gpW. (B) The typlical sigmoidal shape unfolding with steps. (C) Here gpW unfolds via a single step. (D) The gpW protein shows a typical hopping behavior during unfolding.

All four depicted traces show the complete unfolding of the polyprotein with the gpW

signal and six titin I27 unfolding steps. It has to be mentioned here, that the force ramp experiments at this low speed with the titin I27 domain were not reported before using the AFM.

From all measured traces (n=86) the gpW unfolding pattern was again divided into three described groups. N=28 traces were detected in which the gpW unfolds smoothly via a sigmidal shape unfolding curve (panel A). Here the noise during the sigmoidal shape transition can be between 3-5nm. A similar amount of traces however has been found to be in the second cluster (n=25), in which the gpW unfolds via several steps within the sigmoidal shape unfolding signal (panel B). Also here a few traces were detected in which the gpW unfolds via a single step (n=4; panel C). Additionally traces have been observed with a clear hopping signal (panel D, n=29). The amplitude of this hopping signal is with 8-10nm much smaller than the detected total length of the gpW protein (18-20nm). This can be explained by the observation that the hopping signal appears at small forces 1-10pN where the unfolding length of the gpW is much smaller than its total length (see WLC model 2.2).

Figure 7.9 gives an overview of the measured distributions of unfolding force and length of the gpW and for comparison also of the titin I27 domain, as the titin domain has not been measured before at a rate of 1pN/s. Here it has been observed that the unfolding behavior of the I27 domain is still clear two-state even at this slow speed of 1pN/nm near equilibrium. To display the force distribution in a histogram a bin size of 5pN was chosen for the gpW (as in the majority of traces gpW unfolds between 1 and 10pN) and a bin size of 10pN for the titin I27 domain.

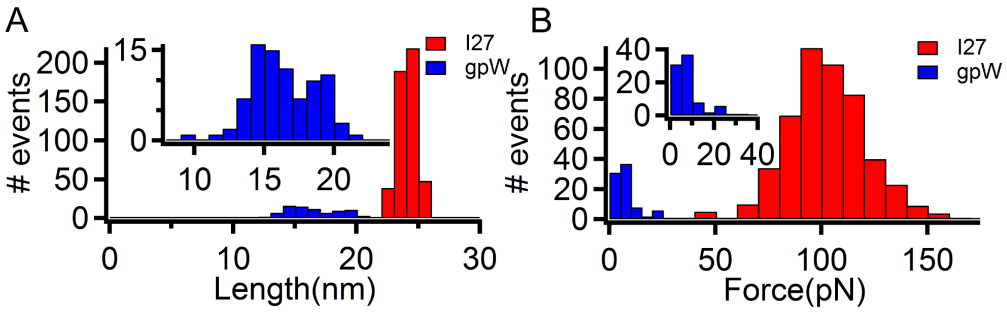

Figure 7.9: Histogram of the distributions of the estimated unfolding length (bin size 1nm) (A) and the measured unfolding force (bin size 5pN for gpW and bin size 10pN for the I27) (B) of the gpW and the titin I27 domain at a constant force ramp of 1pN/s

Averaging the distributions for the gpW results in a mean unfolding length of (16.3 ± 2.3)nm and in a mean unfolding mid force of (7.8 ± 6.7)pN. The average of the distributions of the titin I27 domain are (24.0 ± 0.7)nm and (102.2 ± 18.7)pN.

Here the center of the detected unfolding force distribution of the gpw lies within 5-10pN. Very interestingly during these measurements at 1pN/s traces were detected like in panel D in figure 7.8, in which the gpW seems to hop partially back and forth during unfolding with an amplitude about 8-10nm for 1-2s at a force of 0-5pN.

Summmary and discussion

The Force Ramp measurements on gpW revealed a very distinct unfolding behavior compared to the one observed for the Csp. In the majority of cases gpW was found to unfold under a very small force in the range of 1-20pN, when estimating roughly the force for the midpoint of the sigmoidal shape curve. A shift in this unfolding force was observed from 7.8pN to 13pN to 18pN when increasing the force ramp speed from 1pN/s to 20pN/s to 80pN/s. The following table 7.1 gives an overview of the mid force and length distributions of gpW.

Mid force (pN)	Length (nm)	force rate (pN/s)	Nr of traces
7.8 ± 6.7	16.3 ± 2.3	1	86
14.7 ± 9.5	17.5 ± 2.5	20	78
13.2 ± 13.1	18.3 ± 2.5	40	50
12.7 ± 11.5	18.3 ± 2.5	60	47
18.0 ± 16.2	16.7 ± 2.6	80	24

Table 7.1: Force and length distribution detected for gpW at different force ramp speeds

The total length detected for the gpW coincides well with the expected value, when using the calculated contour length of 20.4nm. When reviewing the structure of gpW (see e.g figure 3.4) its building structure starts from the beginning of the first alpha helix (α_1) and ends at the end of the second alpha helix (α_2), which consists of 51 amino acids. This amount of aminoacids multiplied with 0.4nm (the approximated length of on amino acid) results is 20.4nm. Here it has to keep in mind that the experimental estimated total unfolding length of the gpW is just roughly obtained by measuring the beginning and the end of the sigmoidal shape signal. Therefore the estimated length still could be 1-2nm larger than the true value.

The shape of the unfolding pattern of gpW was found to be sigmoidal and therefore with the exception of a few traces (10-15%) not a clear single or multiple step. Furthermore the sigmoidal shape showed a high degree of fluctuations in length, which amplitude was in the range of 3-5nm.

The population of the three selected groups, describing the unfolding pattern of gpW was changing slightly from the measurements done at 1pN/s to the measurements performed at 20-80pN/s. However a very distinct change could not be observed and all three groups where still populated at each force ramp speed. Additionally the slow rate of 1pN/s at near equilibrium revealed a characteristic hopping pattern with an amplitude of 8-10nm, which is the half of the total estimated length of gpW (16-20nm). However this small value of 8-10nm coincide with the unfolding length of gpW at forces between 1-10pN following the WLC model.

In order to get more insight into the observed hopping behavior of gpW, experiments were conducted in which the protein was hold at 5pN in combining force ramp and force clamp measurements.

7.1.3 Combined constant force and force ramp measurements

Figure 7.10 shows three typical results obtained by holding the gpW protein at a constant force of 5pN for up to 20-30s.

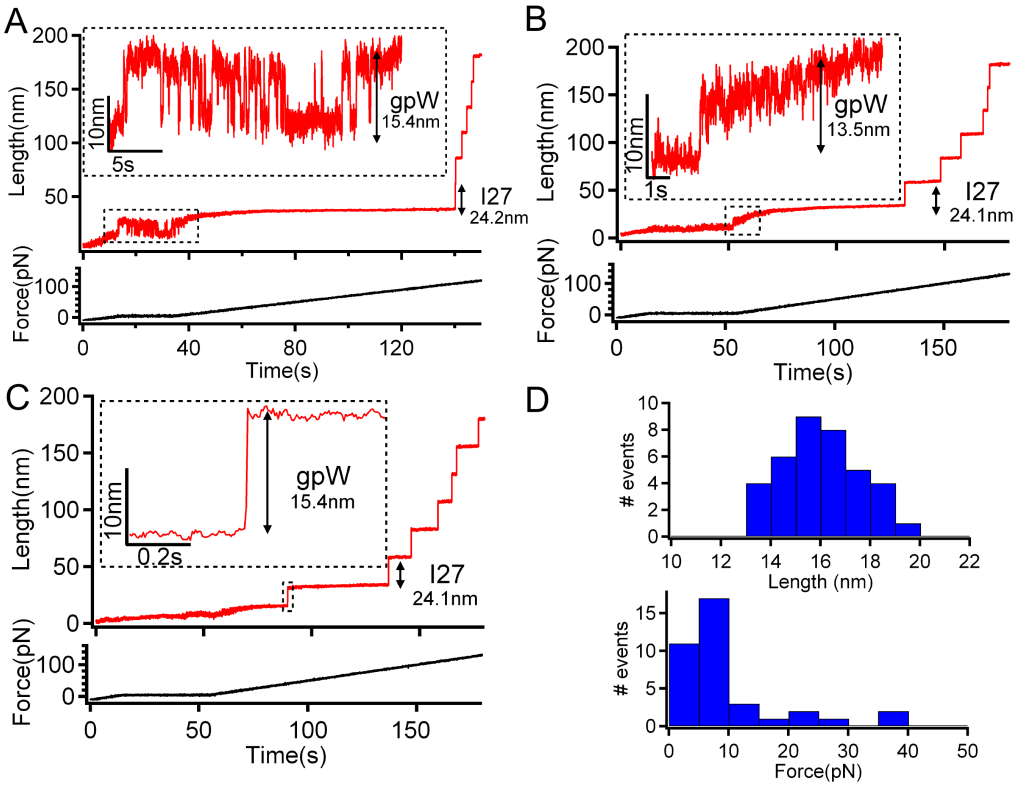

Figure 7.10: Force ramp combined with constant force 5pN traces for gpW at a constant force rate of 1pN/s and with a spring constant of 3-5pN/nm panel A, B and C. Panel D) Distributions of the estimated unfolding length and the measured unfolding force of the gpW.

In order to hold the gpW protein at a constant force of 5pN, the experimental force

sequence consists of starting the force ramp of 1pN/s until 5pN are reached. Then the force was hold constant for 20-30s. Afterwards the force ramp of 1pN/s continues until the titin I27 fingerprint is obtained (approx 150pN). The ramp was started at $|\tilde{F}| = 10$pN pushing force. Hence after the around 15s the constant force of 5pN was reached and applied on the gpW. The majority of traces (n=24) from a total number of (n=37) reveal a hopping signal when the constant force of 5pN is applied (see figure 7.10 panel A). Its amplitude is around 8-10nm and therefore comparable with the hopping amplitude observed at a force ramp speed of 1pN/s (see figure 7.8 panel D). During the performed measurements the hopping signal could be observed up to 30s. As detected during the force ramp measurements at a speed of 1pN/s, the gpW can also here unfold via a sigmoidal shape step (see figure 7.10 panel B, n=8) or via a single step (see figure 7.10 panel C, n=5). The corresponding distributions of gpW (panel D) give an average value of the unfolding force of (9.2 ± 8.7)pN and an unfolding length of (16.0 ± 1.5)nm, which are in the same range like measured at a force ramp speed of 1pNs with the same BioLever B. These distribution of unfolding lengths and force values were just obtained for the corresponding event when the gpW finally unfolds.

Figure 7.11 shows two hopping traces in further detail panel A and panel B.

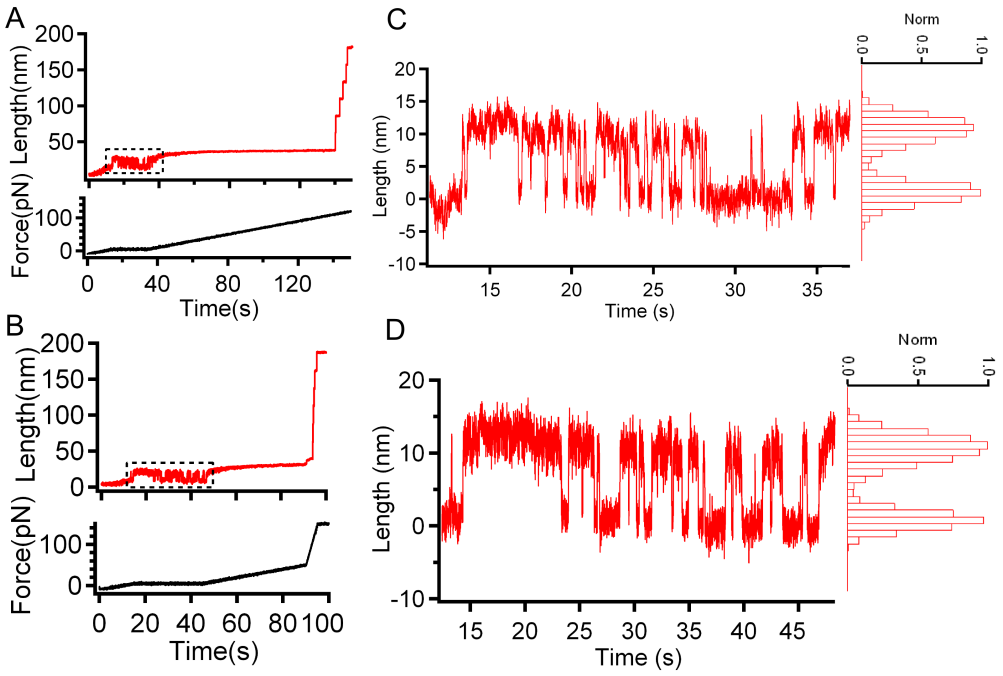

Figure 7.11: Two constant force hopping traces at 5pN for gpW are shown for 20 s (A) and 30s(B). (C) and (D) are showing the corresponding insets from the trajectories seen en (A) and (B) respectively. On the right corresponding the normalized length histogram can be seen

Panel C and panel D are showing the hopping behavior at a constant force of 5pN observed

for gpW for 20 (panel C) and 30 seconds (panel D). In both cases the trajectory was shifted so that the lower state equals a zero length in order to have a better orientation. Additionally the normalized length distributions with 1nm bin size are shown to the right of each trace, which indicate a equal distribution of lengths in both the upper and the lower state. In sum n=11 traces (from n=34) have been selected for the analysis as they showed a long hopping signal between 10 and 30s similar to the two examples shown in the figure 7.11 above. They were used to estimate the length distribution, which is depicted in the following figure 7.12.

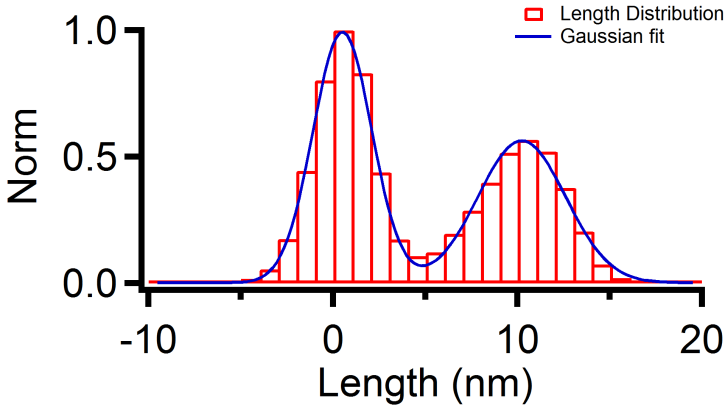

Figure 7.12: Normalized total length distribution of all the eleven selected hopping traces.

The distribution was fitted to a two-peak gaussian. This resulted in an approximate distance of the two peaks of (9.7±0.1)nm. Additionally this distribution reveals that in majority of time within the selected 11 traces the gpW populates the lower state, so in this case the 5pN is still not the equilibrium force $F_{1/2}$.

The hopping signal at a constant force of 5pN now enables to measure the lifetimes in the lower state and in the extended upper state. From these lifetime distribution the corresponding rates can be estimated. As it has been mentioned in the Materials and Methods section before, the build in written procedure function FC_DetectSteps was used to detect the points where the length value changes to the other state.

This analysis resulted in the following length distribution which is shown in figure 7.13. The average of this distribution gives a values of (7.5 ± 1.8)nm, which is by 2nm smaller than estimated above. This difference of 2nm could be explained by the fact that before applying the lifetime analysis function the data was smoothed by a median smoothing factor of ten. However the estimated length distribution is quite broad ranging from 4 to 10nm.

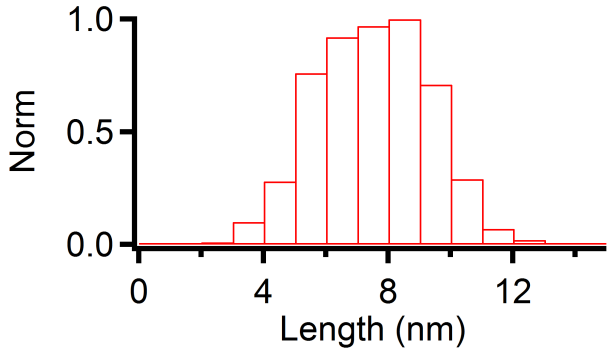

Figure 7.13: Normalized total length distribution of all the eleven selected hopping traces.

The lifetime analysis of the eleven traces resulted in the following time distributions in the lower and in the upper extended state (see figure 7.14).

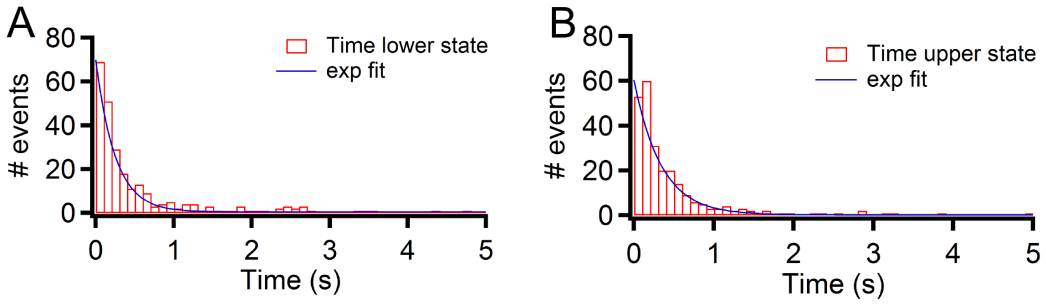

Figure 7.14: Lifetime distributions of the gpW at 5pN constant force in the lower and in the upper state

The exponential fit gives a first approximation of the corresponding rates k_l in the lower and k_u in the upper state of $k_l \approx 4s^{-1}$ and $k_u \approx 3s^{-1}$. This reflects exactly the histogram in figure 7.12 wherein more lengths were counted in the lower state than in the upper extended state. I.e at a force of 5pN more gpW proteins are still folded than mechanically unfolded.

Summary and disscussion

Using a combination of force ramp and constant force measurements enabled a new characterization of the gpW under a mechanical force, which was supposed to be just possible to be achieved with optical or magnetic tweezers. However using the force clamp AFS combined with the BioLever B cantilever gave new insights of the capacity of this instrument. These measurements affirmed the hopping behavior of gpW at a small constant force of 5pN, which has been detected in the force ramp measurements conducted before at a rate of 1pN/s, which can be said to be near equilibrium. The amplitude of the length between the lower and the upper state has been found to lay between 7 and 9nm. The

lifetime analysis of 11 traces revealed rates of the lower and upper state estimated, which are in the range of rate estimated for fast folding proteins under a small constant force. At this point the discussed and already applied stochastic kinetic simulations were performed on gpW.

7.2 Stochastic kinetic simulations on gpW

As the performed stochastic kinetic simulation on the Csp were able to fit the experimental results very well, now they were also performed on gpW again for both cases, for a constant force and for a force changing linearly with time (force ramp). As there are no existing values from the literature for the concrete folding and un(folding) rate under zero force k_f^0 and k_u^0 and neither for the value of the distance to the unfolding transition state Δx_u^0, here the constant force simulations were conducted first in order to estimate from the experimentally determined lifetimes from the constant force measurements the corresponding values under zero force.

7.2.1 Stochastic kinetic simulations of the constant force measurements at 5pN

Using the Bell equation 2.3 an un(folding) rate of $3(4)s^{-1}$ under a force of 5pN results in a very high folding rate under zero force of $k_f^0 \approx 30000s^{-1}$ and in a unfolding rate under zero force $k_u^0 \approx 2.5s^{-1}$, when using a distance to the unfolding transition state of $\Delta x_u \approx 0.2nm$ and a contour length of 24nm like is has been determined for the Csp. A contour length of 24nm was used here as it was fitting the experimental constant force data better than a contour length of 20nm and which is still in the limit of the contour length estimation from the number of amino acids of the gpW. As a constant force of 5pN is very small and small fluctuations in force (deflection) of the cantilever can already have a big impact on the results, an external damping factor of 1/10 needed to be introduced in order keep the simulated result close to the experimentally observed one as it was done for the Csp simulations at a constant force of 20pN (damping factor 1/3).

Figure 7.15 shows four of a total of five simulated length vs time trajectories with the mentioned values used for the (un)folding rates under zero force, the distance to the unfolding transition and the contour length of the gpW. All simulations were calculated for 30s and a time filter of 10ms was applied in order to simulate the experimental sampling rate. Because of the experimentally determined un(folding) rate under force, the gpW is more often in the folded (lower) state than in the extended unfolded (upper) state, which

reflects nicely the good performance of the simulation.

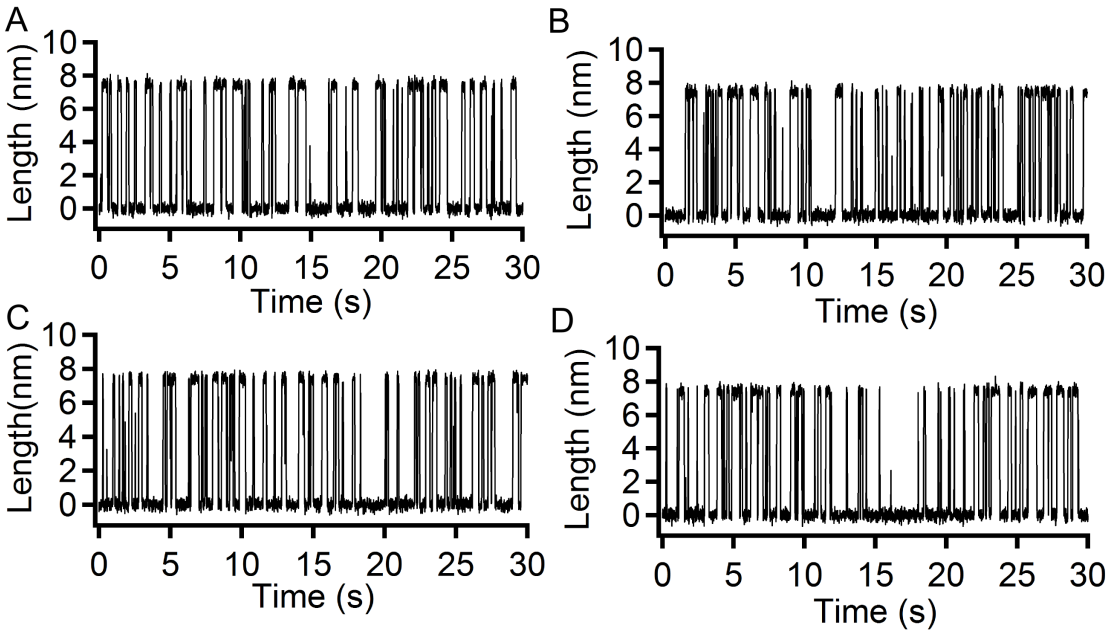

Figure 7.15: Four simulated constant force traces for gpW at 5pN using the discussed stochastic kinetic simulations (A)-(D)

The resulting length histogram (bin size 0.2nm) of the five simulated traces is shown in figure 7.16, which revealed a good agreement of the corresponding length histogram from the eleven experimentally measured traces (see figure 7.12).

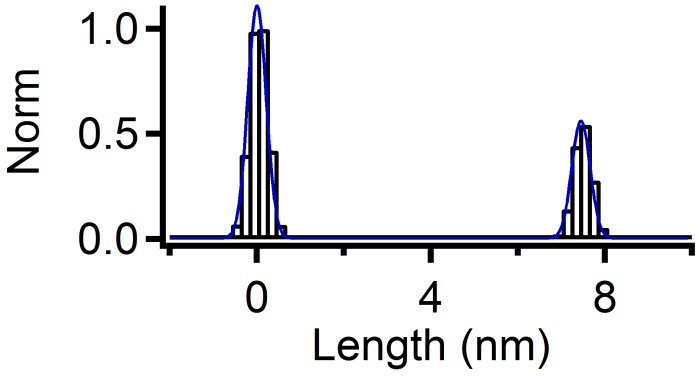

Figure 7.16: Normalized length histogram (bin size 0.2nm) of all five simulated traces

As it can be seen here by using a simple two state model for the simulations the resulting length distribution is very sharp. This is because many factors are not included into the simulation like a possible multi-state behavior of the protein or experimental factors like drift caused for example by the viscous drag of the cantilever. Each peak of the distribution was then fitted to a gaussian and resulted in a inter-peak distance of around (7.43 ± 0.01)nm, which is a somewhat smaller value than determined experimentally in

figure 7.12 (9.7±0.1)nm but which is in a good agreement with the measured length from the lifetime analysis in figure 7.13 (7.5 ± 1.8). For that reason a total contour length value of gpW was chosen to be 24nm as discussed before.

Figure 7.17 shows the comparison of two simulated traces (see figure 7.15 (B) and (D)) to the experimental traces from figure 7.11. Therein both traces were not overlapped, instead the experimental traces was shifted by 15nm in y-direction for a better comparison.

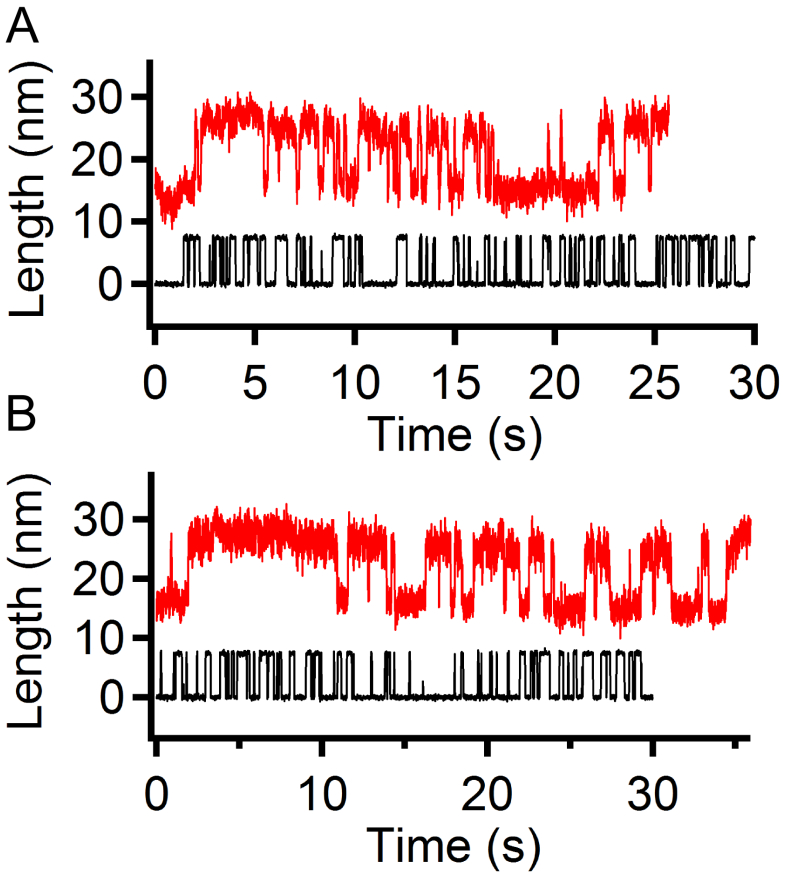

Figure 7.17: Two simulated constant force traces for gpW at 5pN are compared to the experimental trajectories (A)-(B)

As it can be seen here, the noise (fluctuations) of the experimental traces is much higher than compared to the one from the simulations, which caused the discrepancy between the corresponding length values. This could mean that the effect of the cantilever oscillations is still higher than used in the simulations. However the rate being either in the folded or in the extended unfolded state is in good agreement compared to the experimental data, as it was shown in the length histogram of the simulation before (see figure 7.16).

As the results of the simulations showed the expected good agreement with the experimental data at a constant force of 5pN for the gpW, the same parameter were now used to perform force ramp simulations at a rate of 1pN/s.

7.2.2 Stochastic kinetic simulations of the fore ramp measurements at a rate of 1pN/s

Five force ramp simulations at 1pN/s were performed for three different values of Δx_u respectively. Here all figures displaying the conducted simulations were shifted in time to start at 10s in order to have a better comparison later to the experimental traces, in which the force reaches 0pN at a time of 10s. I.e in the following figures an unfolding of gpW after 15s corresponds to an unfolding force of 5pN. Figure 7.18 shows three simulated length vs time trajectories using a Δx_u of 0.2nm.

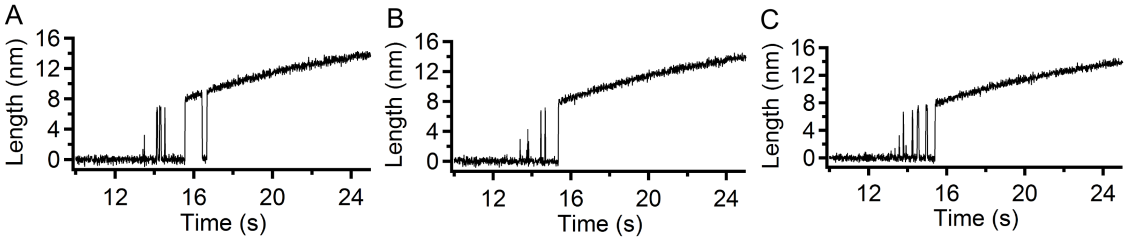

Figure 7.18: Three simulated force ramp traces for gpW at 1pN/s are shown, using a value of 0.2nm for Δx_u (A)-(C)

Here the simulated force ramp traces appear similar to a step unfolding, but with some hopping events before, the unfolding length reaches like 8nm at a final unfolding force of around 5-6pN.

When increasing the Δx_u to 1nm, the fluctuations increase in the simulated traces but the time range of observable hopping events is decreasing, which can be seen in figure 7.19. However, the observed range in force and unfolding length is quite similar than observed before in figure 7.18.

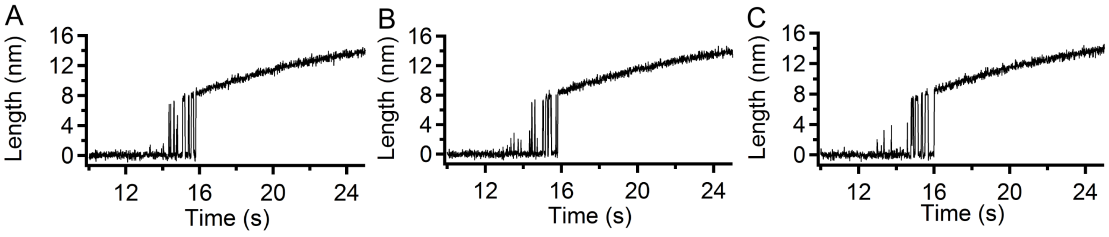

Figure 7.19: Three simulated force ramp traces for gpW at 1pN/s are shown, using a value of 1nm for Δx_u (A)-(C)

When increasing the Δx_u further to 4nm, the fluctuations increase but the time range of unfolding is decreasing so much that the unfolding pattern becomes sigmoidal, which can be seen in figure 7.19. Here the unfolding force is decreasing slightly to values between 4 and 5pN.

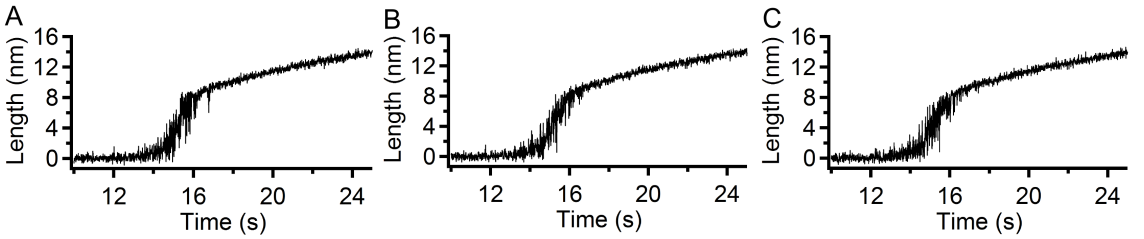

Figure 7.20: Three simulated force ramp traces for gpW at 1pN/s are shown, using a value of 4nm for Δx_u (A)-(C)

Thus by using the same parameter as for the constant force simulations and by changing the Δx_u value, which affects both the folding and unfolding rate under force (see Bell equation 2.3) different unfolding patterns can be generated. Interestingly, all the three shown patterns can be actually found in the force ramp experimental data of gpW at a force ramp of 1pN/s.

Figure 7.21 shows the comparison from each type of the three simulations with a similar detected experimental unfolding trajectory of gpW.

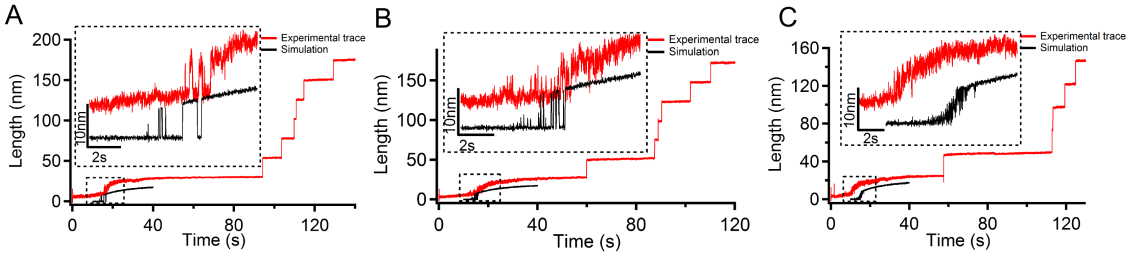

Figure 7.21: Comparison of the experimental force ramp trajectories to the simulated trajectories at a force ramp of 1pN/s (A) Comparison to a simulated trajectory with $\Delta x_u = 0.2nm$ (B) Comparison to a simulated trajectory with $\Delta x_u = 1nm$ (C) Comparison to a simulated trajectory with $\Delta x_u = 4nm$

The simulated trajectories in panel (A) and panel (B) fit quite well the experimental trace shown in red. In panel C the simulated trace is a bit shifted to higher force values than the experimental trace. Here as discussed for the force ramp simulations of the Csp, the difference in height between the experimental and the simulated trace is mainly caused by the linear increase in length by time as the force is controlled to increase linearly. However, in order to demonstrate that the shape of the simulated traces fit very well to the experimental ones (see figure 7.22), the traces from figure 7.21 (A)-(C) were now shifted manually and therefore overlapped.

Thus very surprisingly the simple simulations can give an explanation for the observed heterogeneity of unfolding patterns of gpW during the force ramp measurements. This could be also the explanation why not in all cases a hopping pattern of gpW is detected

during the constant force experiment of 5pN. Although it is not known if the estimated values of k_u^0, k_f^0 and especially the suggested value of Δx_u are that correct (a Δx_u of 4nm is very improbable also because of an existing entropic barrier when extending the protein), still the simulation gives an important hint to solve the observed complexity for gpW. That is, when constraining a fast folding protein like the gpW under a mechanical force a small uncertainty in Δx_u can already shift the unfolding pattern to a very different observation, whereas the values of the un(folding) rates under zero force keep constant. Furthermore this uncertainty should be larger when the energetic folding barrier becomes marginal.

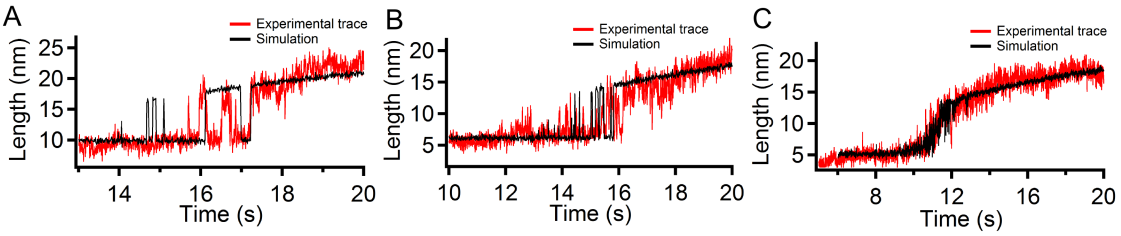

Figure 7.22: Overlap of the experimental force ramp trajectories to the simulated trajectories at a force ramp of 1pN/s (A) $\Delta x_u = 0.2nm$ (B) $\Delta x_u = 1nm$ (C) $\Delta x_u = 4nm$

Thus the simulation can offer a simple and different explanation for the observed heterogeneity when unfolding gpW than the typical one that the gpW sample itself for example is not really pure and consists of different types of medium folded or unfolded proteins or that the AFM cantilever applied a mechanical force on different positions of the sample, which could effect its unfolding behavior.

7.3 SMD simulations

All conducted SMD simulations started from the same MD equilibrium simulation. Therefore the first section will provide the MD results. Then the constant velocity and constant force SMD simulations will be given. The 3Dim structure of gpW is shown in figure 7.23 illustrating the 2 beta strands β_1 and β_2, the two alpha helices α_1 and α_2 and the N and C terminus to have a better orientation for the following sections.

7.3.1 MD equilibrium simulation

The MD simulation was performed on the pdb file 2L6Q.pdb, which was slightly changed by adding three more residues to the C-terminus (PAG). Additionally to the N-terminus five glycine residues were added to really simulate the experimental used structure. Finally,

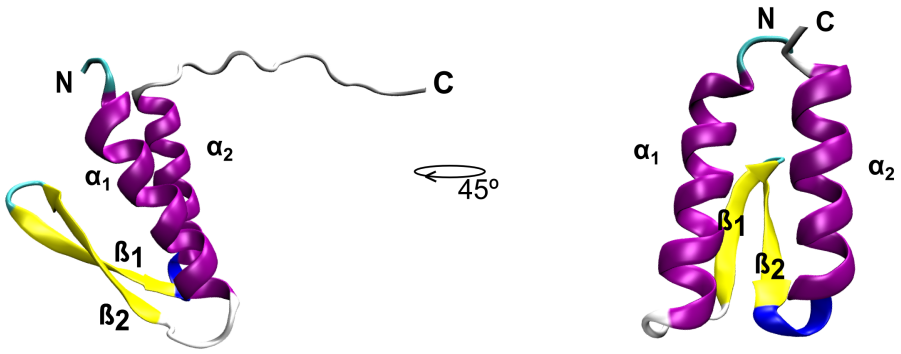

Figure 7.23: The 3Dim structure of the gpW is shown and the structural elements are indicated.

the waterbox was constructed with the dimension (x y z) of (0 0 0)(30 8 7)nm and contained 161315 atoms. Then the gpW inside the waterbox was minimized using steepest decent for 0.2ns and equilibrated for 1ns.

After the simulation was run, the RMSD value was calculated for each timestep of the equilibrium simulation, which compares the position of the backbone atoms of the original and the simulated structure and which is shown in figure 7.24.

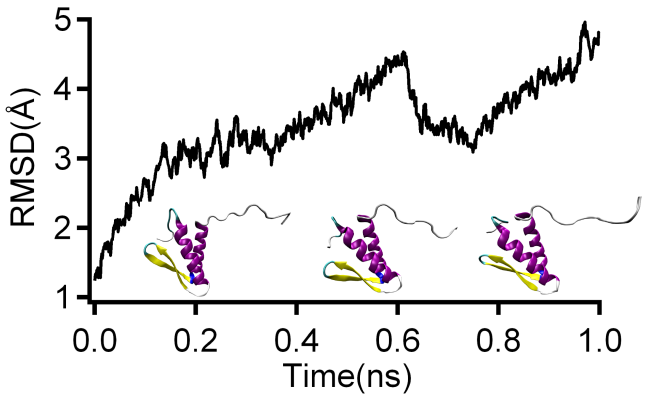

Figure 7.24: Change in the RMSD value (in angstrom) during the MD simulation. Corresponding snapshots of the designed gpW structure at 0.2,0.6 and 0.9ns are shown.

In comparison to the RMSD of the Csp, here the RMSD fluctuates a bit, which is caused basically by the movement of long tails. All conducted SMD simulations started then with the same coordinate file from the MD simulation.

7.3.2 Constant velocity SMD simulations

One SMD force/extension curve at a constant velocity of 1m/s is shown in figure 7.25. Although one simulation is not enough to do statistical predictions of the gpW computational unfolding behavior, it can still reveal a basic idea. Also because the velocity of 1m/s is still 10^7 times faster than the velocities used in the corresponding AFM experiments

(400nm/s) more SMD simulations at a constant velocity were not conducted. The force extension curve panel A in figure 7.25 plots the force applied on the C_α-atom of the C-terminal residue and its change in x-direction during the application of a movement with a constant velocity in x-direction. Panel B shows snapshots of the corresponding trajectory at the given point numbered in panel A from 1 to 5.

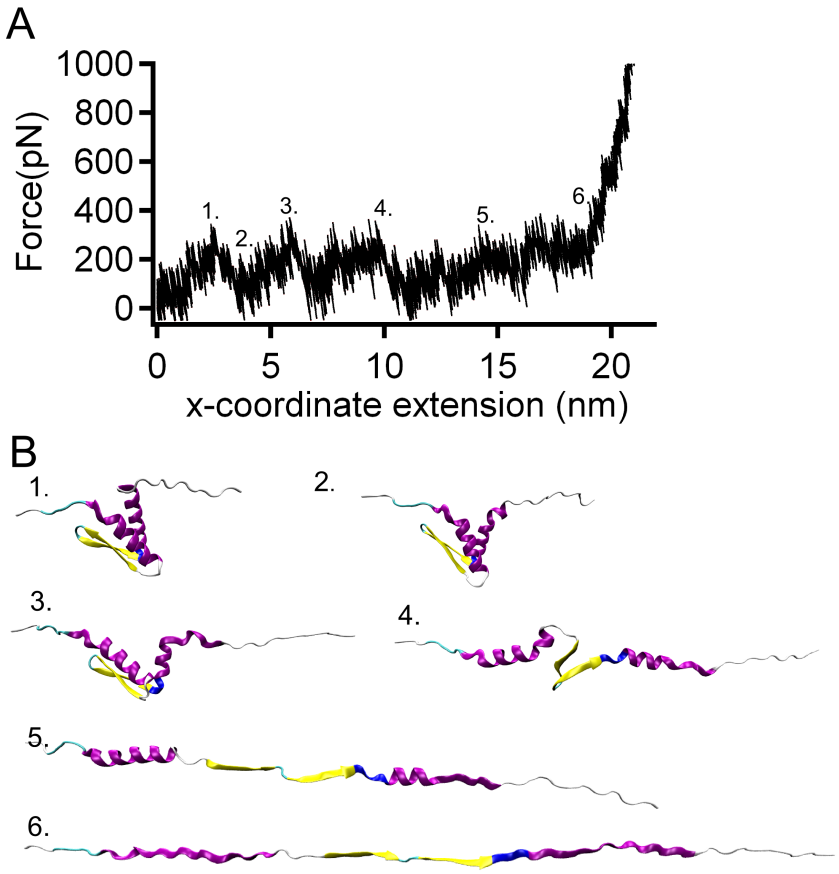

Figure 7.25: A) SMD simulation at a constant velocity of 1m/s of gpW (B) Corresponding snapshots of the trajectory

The x-coordinate extension of the pulling C_α atom was set to zero in the force/extension curve, when the SMD simulation starts in order to have a better comparison to the experimentally estimated length. Compared to the SMD constant velocity simulations conducted with the Csp (see figure 6.38), the unfolding force is lower and reaches 200-300pN. No certain force bearing peak is visible, moreover the unfolding force seems to be regularly distributed throughout the whole unfolding process. This could indicate the continuous unfolding process of disentangling the two alpha helices α_1 and α_2.

The snapshots of the corresponding gpW unfolding trajectory reveal an unzipping of the two helices α_1 and α_2 from point 1 (2.5ns, 2.2nm) to point 3 (6.5ns, 6nm). At point 4 (10ns; 9.6nm) the two beta strands β_1 and β_2 get fully stretched until point 5 (15ns;

14.5nm). From this point on the alpha helices α_1 and α_2 become disentangled, with α_2 disentangling first until the protein is totally stretched at point 6 (20ns, 19.4nm).

7.3.3 Constant force SMD simulations

Figure 7.26 shows all the gpW separation length vs time trajectories from the conducted constant force SMD simulations.

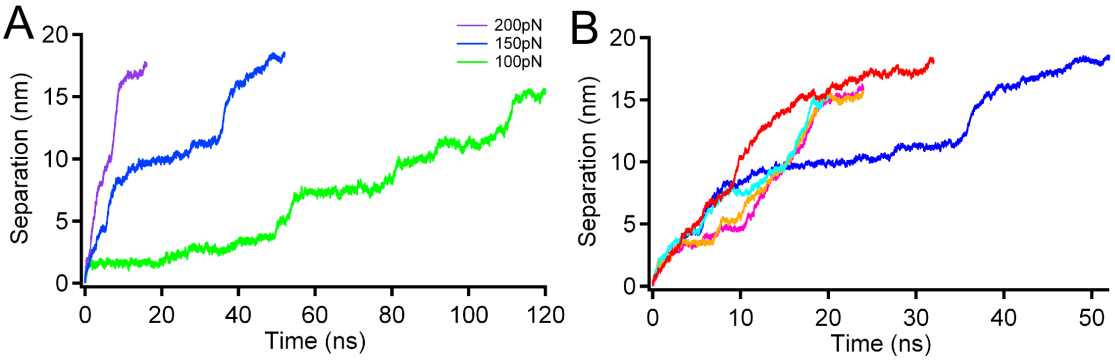

Figure 7.26: (A) Constant force SMD simualtions at 100, 150 and 200pN (B) Five constant force SMD simulations at 150pN

In panel A three trajectories from SMS simulations at different constant forces 100, 150 and 300pN are compared. Here it has been found that 150pN is an ideal force to conduct SMD simulations on gpW, as it unfolds in a reasonable calculation time, which can be performed by the used computational resources (see blue trace). Panel B shows the variety of five separation length vs time traces, all conducted at a constant force of 150pN. All gpW unfolding trajectories show a very similar pathway. In general first the two alpha helices α_1 and α_2 get separated and than it depends if first alpha helix α_2 totally unravels or if the two beta strands β_1 and β_2 get totally stretched. These two cases will be shown in the following figures.

Figure 7.27 illustrates the details of the blue trajectory from figure 7.26 (A) and (B). Panel A shows the single separation vs time trajectory, which indicates a two step unfolding behavior, the first until 10ns and the second around 35ns. The overall unfolding time is 52ns. Panel B shows the corresponding secondary structure content plotted using the timeline plugin of the VMD program. Together with the snapshots shown in panel C, the whole unfolding trajectory can be followed. The unfolding pathway of this trajectory is very similar to the one observed during the constant velocity simulation (see figure 7.25 panel B) The unfolding starts by opening the two alpha helices α_1 and α_2, which happens between 1.5 and 2ns. Then the two helices get further separated until until they are totally in parallel to the pulling direction at around 10ns.

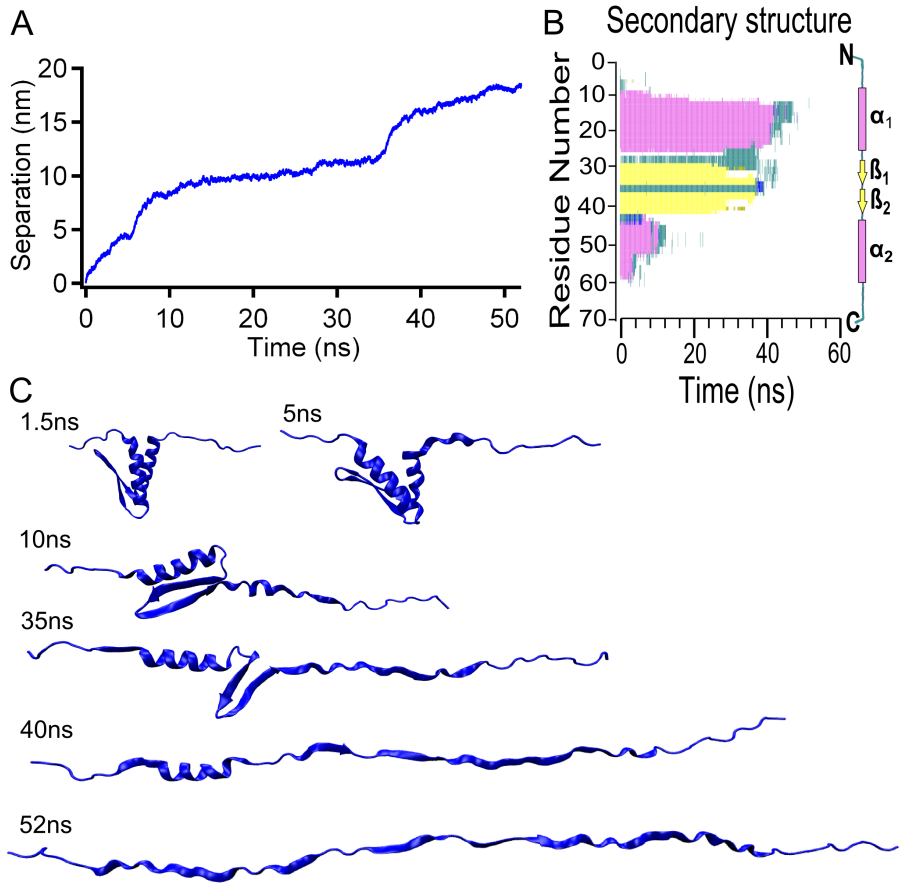

Figure 7.27: (A) Constant force SMD simualtions at 15pN (B) Secondary structure content (C) Snapshots from the trajectory of the gpW, taken at the indicated time.

At this point the alpha helix α_2 becomes totally stretched and the gpW stays in this state until a time of 35ns, when the two beta strands β_1 and β_2 get detached from the alpha helix α_1. During the last 5ns the two beta strand than get totally stretched at the time of 40ns. Afterwards the alpha helix α_1 gets unraveled until the gpW protein is totally stretched at 52ns. In this simulation the gpW stays at the intermediate around 25ns in which the two beta strands are still interacting with the alpha helix α_1.

For comparison figure 7.28 illustrates the details of the red trajectory from figure 7.26 (B). Panel A shows the single separation vs time trajectory, which does not indicate any clear distinguishable intermediate steps. Here the unfolding pattern of the gpW has a quite linear shape. The overall unfolding time is 32ns. Panel B and C again facilitate to follow the whole unfolding trajectory. Although the unfolding starts like in the red trace in figure 7.27 by opening the two alpha helices α_1 and α_2, which happens at a time of around 2ns, the unfolding pathway of this trajectory is a bit different. Here first the two beta strands β_1 and β_2 get detached from the alpha helix α_1, which happens around 6ns. Then the beta strands β_1 and β_2 are totally elongated further until a time of around 13ns. Afterwards

the two alpha helices get totally stretched until 32ns with the α_2 helix unfolding first.

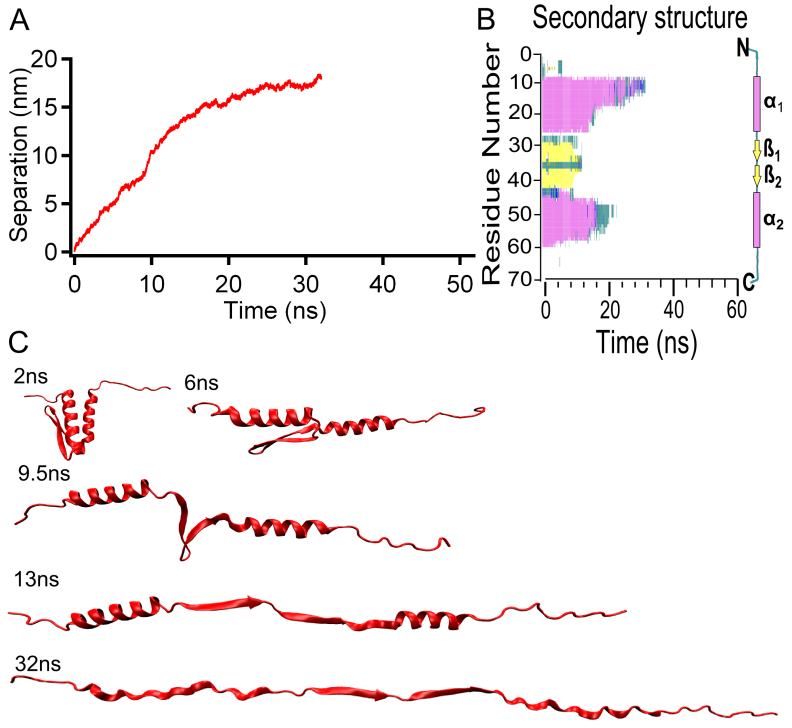

Figure 7.28: (A) Constant force SMD simualtions at 150pN (B) Secondary structure content (C) Snapshots from the trajectory of the gpW, taken at the indicated time.

Figure 7.29 compares the two discussed unfolding trajectories (panel A).

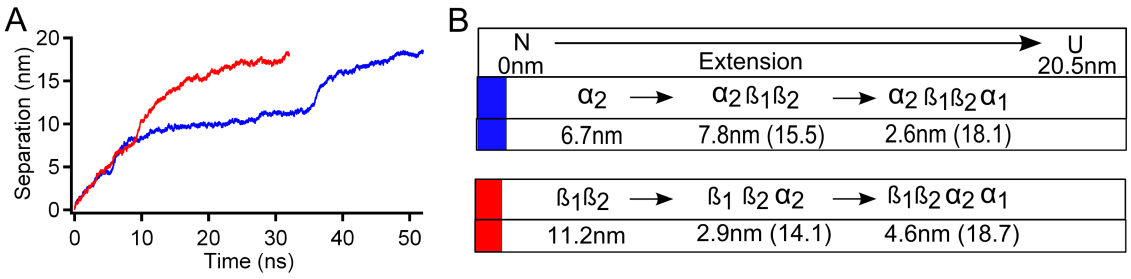

Figure 7.29: (A) Constant force SMD simualtions at 150pN (B) Length distributions during the detected unfolding pathway

The table in panel B, shows the length distributions detected for each of the two unfolding pathways. It indicates not only the single step lengths, furthermore also the absolute lengths are shown in brackets.

Summary and Discussion

The conducted SMD simulations describe the unfolding pathway of the gpW at atomic resolution. Therein a simple unfolding pathways has been identified, which is just slightly

varying. All detected unfolding pathways at a constant force start with the unzipping of the two alpha helices α_1 and α_2 until both helices are parallel to the pulling direction. The next step is the the loss of interaction between the two beta strands β_1 and β_2 to the alpha helix α_1. Then the two beta strands get disrupted and totally stretched. The detected pathways can slightly differ depending on when the alpha helix α_2 unravels. The blue and red curve from figure 7.29 corresponds to this difference found in the unfolding pathways of gpW.

While during the red trajectory both alpha helices get disentangled after the disrupting of the two beta strands during the blue trace the alpha helix α_2 disentangles first before the two beta strands get disattached. Here it seems that the force of disentangling the alpha helix is comparable to the force needed to disrupt a anti-parallel beta sheet.

In sum, the performed SMD simulations cannot really help to interpret the experimentally observed complexity of the unfolding of gpW. This is also because the timestep (fs) and the unfolding time of the trajectory (ns) are much below the corresponding

7.4 Conclusions

The mechanical unfolding mechanism of gpW was found to be much more heterogeneous than the one of the Csp. Furthermore when using the force-clamp AFM the described gpW unfolding behavior became much more complex, indicating many different shapes of unfolding patterns ranging from a sigmoidal shape over a combination of sigmoidal shape with single steps to a single step unfolding pattern. Moreover at a force ramp rate of 1pN/s a hopping behavior of gpW became visible. Within these experiments the stable measuring capacity of the force-clamp AFM was reached by applying such a small rate near equilibrium and could just be performed in using the BioLever cantilever with its small spring constant of 5pN/nm. After this discovery constant force measurements at a force of 5pN were successfully conducted with the AFM, which were not reported before by any other single molecule force spectroscopic experiment using the AFM. So far measuring proteins at a low force range, in which hopping of a protein can be observed were only discovered with the AFM in constant velocity mode with a velocity close to equilibrium [148] or by using optical tweezers, because of their low spring constants [93, 141].

Just because of the in this work discovered capacity of the force-clamnp AFM in the very low force regime, the gpW could be investigated in detail and its unfolding force range of 1-20pN could be determined from the force ramp measurements at rates from 1 (1-10pN) to 80pN/s (1-20pN) (see table 7.1). It has to be kept in mind that the estimated force in

table 7.1 is just a very rough approximation, wherein the unfolding force was measured in the middle of the unfolding pattern of gpW. The conducted lifetime analysis of the constant force traces at 5pN revealed un(folding) rates in the range of s^{-1} which coincides well with a protein, which folds under zero force in the range of μs and therefore coincides with the approximated gpW kinetic data from the T-jump relaxation times [78]. Here a concrete determination of the rates is not possible, because the distance to the unfolding transition state is not known and can just be suggested. E.g a value of 0.2nm for Δx_u resulted in a folding rate of k_f^0 of $30000s^{-1}$ when suggesting a contour length of 24nm.
However the complexity of the measurements of gpW resulted from the high variety of observed unfolding patterns detected with the force ramp and constant force measurements. The discovered patterns revealed a huge discrepancy from the usually observed single or multiple unfolding steps. Instead here sigmoidal shape unfolding patterns and combination of sigmoidal shape and steps became apparent, which were also observed here for the first time during force-clamp AFM measurements. As a result a accurate determination of the unfolding force and length of gpW became very difficult within the force ramp measurements and only rough approximations were be realized.
Furthermore the observed unfolding patterns were tried to distribute into three different groups (sigmoidal shape, sigmoidal shape with steps and single steps). However even this classification is difficult to realize and therefore a certain trend of a change in the probability for gpW to unfold via one or another pattern under a different force ramp speed cannot be stated here. Instead it seems, although the detected unfolding force of gpW changes slightly from 1-10pN (at a force ramp of 1pN/s) more to 1-20pN (at a force ramp of 20-80pN/s), that at all force ramp speeds still all groups of unfolding patterns keep being populated with only minor changes.
At this point one first interpretation, which could explain the different observed unfolding patterns is that the gpW sample itself is very heterogeneous because of technical issues related to the purification process therefore having a certain percentage of partially folded proteins, which could show different unfolding behaviors. Another interpretation could come from the AFM experimental setup itself, wherein for example the AFM cantilever picks up the protein at a different point or the protein is sometimes already denaturated on the gold surface. However although both interpretations cannot be totally excluded, similar observations when measuring the Csp sample were not occurring. Therein the sample was always unfolding via single or multiple steps. This concludes that the experimental setup and the purification technique used are appropriate for performing SMFS

measurements.

Another interpretation for the origin of the observed complexity of the unfolding behavior of gpW came from the performed stochastic kinetic simulations. With the help of the performed constant force measurements values for the rates k_f^0 and k_u^0 could be approximated and used for the force ramp simulations. Therein a visible change in unfolding pattern of gpW from a single step to a sigmoidal shape unfolding curve was observed by changing the Δx_u value (see figure 7.22). I.e. an uncertainty in the Δx_u could also cause different unfolding patterns of the gpW sample. The origin of this explanation comes from the idea that for a protein like gpW with a marginal energy barrier, the uncertainty of Δx_u should be larger than for proteins with a higher energy barrier like Csp or the titin I27 domain. That is why the conducted stochastic simulations were very helpful to gain another explanation for the observed complexity of the unfolding behavior of gpW. This explanation is further strengthened with the fact that different unfolding pathways can already coexist in a two-state folding protein sample [203] and furthermore can be revealed directly by SMFS experiments using the AFM, as it has been observed in the chapter before when measuring the Csp sample.

Thus the mechanical unfolding behavior of gpW can be described as highly heterogeneous showing different unfolding pathways within a unfolding force range of 1-20pN. The impact of the fast folding rate on the unfolding pattern could be experimentally observed with the force-clamp AFM measurements and computationally derived from the stochastic kinetic simulations. This allowed to conclude that first gpW folds in a μs range and furthermore that the folding mechanism of gpW seems to be more complex than a simple two-state mechanism, as not all unfolding patterns especially like the ones with a sigmoidal shape including steps could be represented with the two-state stochastic stimulation.

Interestingly, not only in the case of the gpW as a fast folding protein also when measuring the BBL protein with an AFM force ramp experiment different unfolding patterns could be detected being very similar to the observations from the gpW experiment (see appendix 2 V). This could mean that beside gpW more fast folding proteins could exist showing this heterogeneity and complexity under a mechanical force, which could be directly observed with the force-clamp AFM.

Part IV

Final conclusions and future perspectives

Conclusions

Single-molecule force spectroscopy allows the application of a mechanical force to the protein, enabling to follow the protein (un)folding reaction within a well-defined single reaction coordinate, the direction of the applied force. Although an impressive improvement of the performance of single molecule experiments (using fluorescence or mechanical force) with respect to the time resolution has been observed during the last decade, single molecule protein folding experiments have almost invariably produced simple two-state observations, which restricts a detailed investigation of the protein un(folding) mechanism. In contrast, here in this work single molecule force experiments on the three proteins of investigation Csp, gpW and also BBL unveiled a high heterogeneity and complexity in their mechanical unfolding behavior, when being investigated in the low force regime with the force-clamp AFM.

The heterogeneity of Csp is manifested as a manifold of individual behaviors that ranges from single step unfolding to events that consist of multiple intermediates, some of which remain stably formed for many seconds during the unfolding experiment. Additionally the probability for Csp to unfold via multiple steps could be modulated by the applied force reaching its highest diversity at a constant force range between 40-60pN. No connection has been found between the order of observed unfolding steps, their individual lengths and measured unfolding force. This heterogeneity was further confirmed by the performed cluster analysis. The conducted SMD simulations showed a very good agreement to the experimental results, that the Csp can unfold via one or multiple steps and that it can also unfold via different pathways. However a correlation between the unfolding lengths of secondary structure elements of Csp to the experimentally observed length distribution could not be established. This discovered unfolding behavior of Csp under force is in stark contrast to other single molecule measurements, wherein Csp has robustly shown simple two-state-like (un)folding when studied with single molecule resolution using both fluorescence detection and force extension AFM in the higher force regime. Hence SMFS measurements on Csp reveal that the pulling force seems to have two effects on the folding

free energy landscape (FEL). First, the mechanical force tilts the FEL to the extended unfolded state and therefore increases the unfolding kinetics, which is a general observation for proteins. Second, the pulling force increases the roughness of the FEL that increases transient trapping and therefore intermediate states become visible during the mechanical unfolding of Csp.

Investigations on the gpW revealed an even much higher degree of heterogeneity and complexity. Here the discovered capacity of the force-clamp AFM to perform stable measurements at a very slow force rate of 1pN/s near equilibrium and at a constant force of just 5pN enabled a detailed mechanical description of the gpW unfolding behavior under a mechanical force. Both types of measurements (force ramp and constant force) at this low speed and force regime using the force-clamp AFM have not been reported before. Therein gpW was found to unfold at very low forces between 1-20pN and a characteristic hopping signal became apparent during the constant force measurements at 5pN. The additional complexity of gpW evolved from the high variety of unfolding patterns discovered during the force ramp measurements. This complexity could be partially resolved with the developed stochastic kinetic model, which showed that the observed unfolding pattern of gpW could be caused by its high folding rate in the μs range and by an existing uncertainty in the distance to the unfolding transition state Δx_u, which is expected for proteins folding over a marginal small energetic barrier. By changing this parameter of Δx_u the stochastic kinetic simulation was able to reproduce the observed sigmoidal unfolding patterns of gpW.

Indications of the mentioned heterogeneity and complexity in the mechanical unfolding behavior of gpW was also found for the BBL domain using constant velocity measurements with different polyprotein constructs. Additionally, measurements on the polyprotein constructs $I27_2$-BBL-I27-BBL-I27-BBL-$I27_2$ and $pFS-BBL_2$ sample unveiled an that multiple BBL domains are not unfolding consecutively one after another, instead they were observed to produce an even higher degree of complexity in the resulting unfolding patterns.

In sum, this work declares that single molecule force spectroscopy measurements with the force-clamp AFM can indeed discover existing complexity in the unfolding behavior of fast folding proteins. Additionally this work showed that a mechanical unfolding behavior of a protein cannot be derived from chemical denaturant unfolding experiments. Another important discovery was that the application of a simple stochastic kinetic simulation on the one hand and atomistic SMD simulations on the other hand confirmed their strength

in interpreting the experimental data. Thus, force-clamp AFM in combination with simulations (stochastic kinetic and SMD) has been shown to emerge as a powerful tool for probing at high resolution the topographic features of the free energy landscapes of two-state and fast folding proteins and how these proteins respond to a force perturbation. Probing these mechanical features of proteins can improve the general understanding of protein folding.

Future perspectives

The discovered mechanical unfolding behavior of Csp can now be used as a prototype for developing simulation and kinetic models in order to describe its multi-state unfolding behavior, as it can be directly detected by using single molecule force spectroscopic techniques. Furthermore constant force measurements at a lower force than 20pN should also reveal a hopping behavior of this protein as it has a fast folding rate of $550s^{-1}$. Therein it would be interesting to observe if Csp also refold through multiple steps. I.e if it can partially refold from one intermediate state to another or if it refold through a single step. As the Csp is described to bind to RNA as it contains a RNA binding motif, also here very interesting experiments could be though of, if for example the binding to RNA will alter its mechanical stability. Another interesting experiment would be the effect of a change in temperature, as cold-shock proteins are described to be expressed under low temperatures maybe a temperature dependence of its mechanical stability could be found, which changes its unfolding behavior.

The findings of the gpW now allow to investigate its folding behavior in detail, in order to detect if it is folding in a strict two-state manner or more with a partial downhill description. Furthermore constant force measurements at different forces should enable to find the force $F_{1/2}$ at which the protein is equally distributed in the lower and in the upper state to estimate its free energy.

Force clamp experiments on the BBL should show very similar findings to the performed measurements on the gpW. These measurements could help to investigate if the pH-value dependence of BBL would alter its mechanical stability and therefore its unfolding behavior.

Conclusiones

La espectroscopía de fuerza de molécula individual (EFMI) permite ejercer fuerzas mecánicas en los extremos de una única proteína. Esto hace posible observar el (des)plegamiento de la proteína a lo largo de una coordenada de reacción bien definida, que es la extensión molecular proyectada en la eje de la fuerza aplicada. A pesar de los grandes avances técnicos logrados a lo largo de la última década en el campo de los experimentos de molécula individual (tanto en fluorescencia como en espectroscopía de fuerza) para mejorar la resolución temporal, los resultados experimentales producen casi siempre señales dicótomas. Así, es muy difícil estudiar el mecanismo que usan las proteínas para plegarse i desplegarse más allá del simple modelo de dos estados (plegado-desplegado).

En esta tesis se ha estudiado el comportamiento de tres proteínas –Csp, gpW y BBL– en un régimen de fuerzas bajas, mediante técnicas de espectroscopía de fuerza de molécula individual usando microscopía de fuerza atómica (MFA). Las mejoras introducidas en los distintos sistemas experimentales han permitido optimizar la resolución temporal en los experimentos, lo que ha hecho posible estudiar la gran complejidad y la alta heterogeneidad en el mecanismo de desplegamiento mecánico de las tres proteínas.

La heterogeneidad en el comportamiento de la proteína Csp se ha observado a través de un amplio espectro de patrones experimentales obtenidos durante su desplegamiento mecánico. Algunas veces el desplegamiento muestra un único estado, mientras que otras se observan múltiples estados intermedios con tiempos de vida en ocasiones del orden de segundos. Además, se ha observado que es posible modular la probabilidad de Csp de desplegarse vía estados intermedios cambiando la fuerza aplicada en los extremos de la proteína. La máxima heterogeneidad en el comportamiento se alcanza alrededor de los 40-60 pN. Adicionalmente, se han realizado simulaciones de dinámica molecular dirigida que confirman nuestros resultados experimentales: Csp puede desplegarse vía un único estado o bien mediante múltiples intermedios, siguiendo diferentes caminos durante la reacción. En ningún caso se ha encontrado una conexión entre la extensión molecular de los distintos estados intermedios observados y la fuerza aplicada, así como tampoco

se ha establecido ninguna correlación entre los elementos que constituyen la estructura secundaria de Csp y los de los estados intermedios observados. Estudios previos de espectroscopía de fuerza a altas fuerzas y fluorescencia sugerían un comportamiento de dos estados para el (des)plegamiento de la proteína Csp. Así, los resultados encontrados en el curso de esta tesis doctoral implican que la acción de una fuerza mecánica aplicada a los extremos de Csp tiene dos efectos importantes en su paisaje de energía libre. Primero, la fuerza favorece los estados con mayor extensión molecular, de modo que inclina el paisaje de energía libre hacia el estado desplegado y extendido, incrementando al mismo tiempo la probabilidad de desplegarse en función del tiempo. Dicho comportamiento es general para todas las proteínas. Segundo, la acción de una fuerza en la proteína modifica su perfil del paisaje de energía libre, favoreciendo la aparición de estados estables intermedios entre el plegado y el desplegado.

Los estudios realizados con la proteína gpW sugieren un comportamiento más heterogéneo y complejo que en el caso anterior. En este caso, el estudio del mecanismo de desplegamiento de gpW se ha realizado mediante espectroscopía de fuerza manteniendo la proteína sujeta a una fuerza constante de solo 5 pN, o bien cambiando el valor de la fuerza aplicada casi adiabáticamente a una velocidad de 1 pN/s. En ambos experimentos (fuerza constante y velocidad constante controlada en fuerza) se ha observado que la proteína gpW se despliega entre 1 y 20 pN, así como también se ha obtenido una señal característica de *hopping* en los cambios de extensión molecular alrededor de 5 pN. La gran diversidad de caminos de desplegamiento observados durante los experimentos realizados a velocidad constante sugieren una gran complejidad de mecanismos para la proteína gpW. Modelos cinéticos estocásticos muestran que dicha complejidad puede ser una consecuencia de la alta frecuencia de plegamiento (en el rango de μs) de la proteína junto con la incertidumbre en la distancia Δx_u entre el estado de transición y el estado plegado, propia de proteínas que se pliegan a través de una barrera energética marginal. De hecho, simulaciones de los modelos estocásticos que incorporan la incertidumbre en Δx_u reproducen la señal sigmoidea obtenida en los patrones de desplegamiento de gpW.

La mencionada heterogeneidad y complejidad en el mecanismo de desplegamiento encontrado para gpW también se ha hallado también para la proteína BBL en experimentos llevados a cabo en el modo de velocidad constante con distintos constructos poliporteicos. Medidas en los constructos $\mathrm{I27_2}$-BBL-I27-BBL-I27-BBL-$\mathrm{I27_2}$ y $\mathrm{pFS-BBL_2}$ han mostrado que los múltiples dominios de BBL no se despliegan consecutivamente uno tras otro, sino que se obtiene una complejidad aún mayor en los patrones de desplegamiento.

En resumen, esta tesis doctoral muestra que los experimentos de espectroscopía de fuerza de molécula individual realizados con microscopio de fuerza atómica en el modo a fuerza constante puede ayudar a descifrar la complejidad en los mecanismos de desplegamiento de proteínas con plegamiento ultra-rápido, más allá de lo que hasta ahora experimentos de desnaturalización química han conseguido observar. La combinación de estos experimentos con simulaciones de modelos estocásticos sencillos o simulaciones atomísticas de dinámica molecular más complejas puede ser de gran ayuda para interpretar los resultados. La combinación de la microscopía de fuerza atómica con simulaciones computacionales (estocásticas y de dinámica molecular) es, en conclusión, un enfoque muy potente para investigar el perfil del paisaje de energía libre de proteínas con mecanismos de (des)plegamiento muy diversos, ya sean ultra-rápidas o de dos estados. El estudio de toda esta fenomenología ayudará en la comprensión general del problema del plegamiento molecular.

Perspectivas futuras

El mecanismo de desplegamiento de la proteína Csp estudiado en esta tesis doctoral servirá como prototipo para desarrollar nuevos modelos teóricos que describan su comportamiento multi-estado, tal y como ha sido observado usando el microscopio de fuerza atómica y los experimentos de espectroscopía de fuerza de molécula individual. Experimentos en los que una fuerza constante menor que 20 pN se aplique a los extremos de la proteína deberían mostrar un comportamiento de *hopping*, ya que ésta tiene una probabilidad de plegamiento por unidad de tiempo de 550 s^{-1}. La cuestión de si es posible volver plegar parcialmente la proteína entre sus estados intermedios o bien si el plegamiento sigue un comportamiento de dos estados entre el plegado y el desplegado sigue abierta. Además, la proteína Csp contiene un lugar de unión específico en su estructura nativa para interaccionar con ARN. El estudio del cambio en su estabilidad mecánica en presencia de ARN podría ayudar a descifrar la función de los estados intermedios observados en experimentos de desplegamiento mecánico. Otro experimento interesante sería estudiar el cambio de estabilidad en función de la temperatura; las proteínas *cold-shock* típicamente se expresan a bajas temperaturas. Así pues, el estudio de la estabilidad termo-mecánica de la proteína podría esclarecer la función de los estados intermedios en el camino de reacción.

Los resultados encontrados con la proteína gpW hacen que sea ahora importante estudiar los mecanismos de plegamiento en detalle, para clarificar si se trata de un proceso de dos o más estados, o si más bien una descripción tipo *downhill* se ajusta mejor. Experimentos

a fuerza constante en un rango amplio podrían ser útiles para caracterizar la fuerza $F_{1/2}$ en la que la proteína tiene la misma probabilidad de encontrarse en el estado plegado y en el desplegado, para así poder determinar su energía libre de formación.

Experimentos a velocidad constante con la proteína BBL deberían proporcionar resultados similares a los obtenidos en esta tesis con la proteína gpW. Estas medidas serían de ayuda para investigar cómo la estabilidad mecánica de BBL depende del valor del pH, lo que podría ser también clave para explicar el mecanismo de desplegamiento de la proteína.

Part V

Appendixes

Appendix 1

NAMD File

```
# NAMD Config file - autogenerated by NAMDgui plugin
# Author: Jan Saam,  saam@charite.de

# input
coordinates             TmCSPwlarge.pdb
structure               TmCSPwlarge.psf
parameters              par_all27_prot_na.prm
paratypecharmm          on

# output
set output              TmCSPmeq18122013
outputname              $output
dcdfile                 ${output}.dcd
xstFile                 ${output}.xst
dcdfreq                 1000
xstFreq                 1000

binaryoutput            no
binaryrestart           no
outputEnergies          1000
restartfreq             1000

fixedAtoms              off

# Basic dynamics
exclude                 scaled1-4
1-4scaling              1
COMmotion               no
dielectric              1.0

# Simulation space partitioning
switching               on
switchdist              9
cutoff                  10
pairlistdist            12

# Multiple timestepping
firsttimestep           0
timestep                1
stepspercycle           20
nonbondedFreq           2
fullElectFrequency      4



# Temperature control

set temperature         298
temperature             $temperature;  # initial temperature

# Langevin Dynamics
langevin                on;            # do langevin dynamics
langevinDamping         1;              # damping coefficient (gamma) of 1/ps
langevinTemp            $temperature;   # bath temperature
langevinHydrogen        no;             # don't couple langevin bath to hydrogens
seed                    12345

# Pressure control
langevinPiston          on
langevinPistonTarget    1.01325; # in bar -> 1.01325 bar = 1 atm
langevinPistonPeriod    200
langevinPistonDecay     100
langevinPistonTemp      $temperature
useFlexibleCell         no
useGroupPressure        no
fixedAtomsForces        off

# PBC
cellBasisVector1        300 0 0
cellBasisVector2        0 70 0
cellBasisVector3        0 0 70
cellOrigin              150.3639678955078 35.08951187133789 34.988136291503906
wrapAll                 on
dcdUnitCell             yes

PME                     yes
PMEGridSizeX            300
PMEGridSizeY            72
PMEGridSizeZ            72


# Scripting

minimize            200000
reinitvels          $temperature
run                 1000000
```

Appendix 2

Force ramp measurements on BBL at a rate of 20pN/s

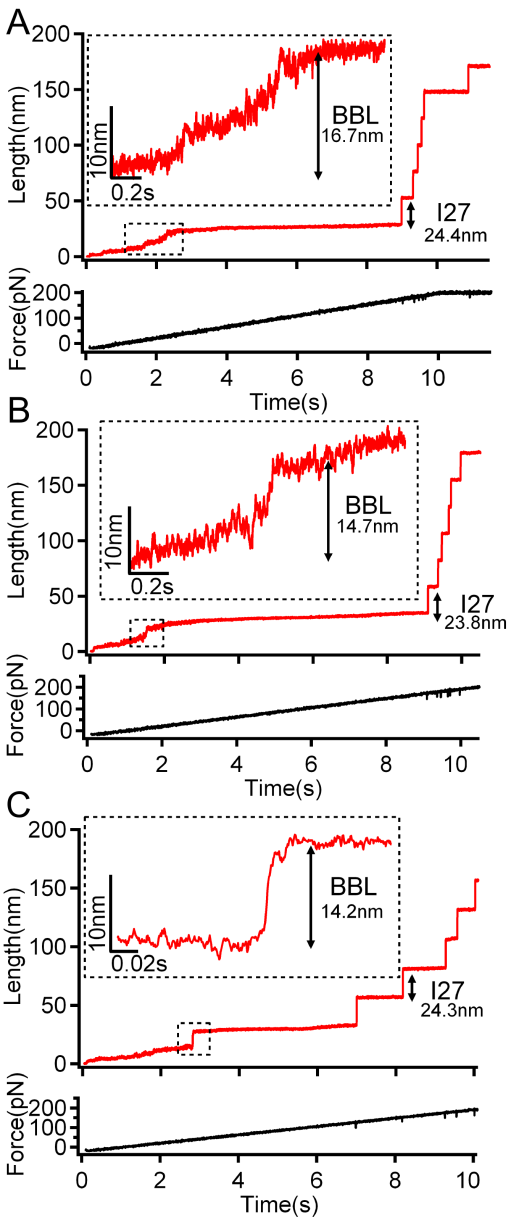

Figure 30: Force ramp traces for the BBL domain at a force ramp rate of 20pN/s with a spring constant of 10-20pN/nm and in PBS buffer pH 7.4

Figure 30 shows three typical force ramp traces at a force ramp rate of 20pN/s, done with the MLCT cantilever D with a spring constant of 10-20pN/nm and in PBS buffer pH 7.4. All three depicted traces show the unfolding of the polyprotein with the BBL signal and five-six titin I27 unfolding steps. All force ramp traces were able to be clustered like the ones of the gpW protein into three clusters. From all measured traces (n=35), n=10 were classified into the first cluster, in which the BBL domain unfolds smoothly via a sigmoidal shape unfolding curve (panel A). Herein the noise during the sigmoidal shape transition can be between 2-3nm and is therefore a bit smaller than it has been detected for the corresponding gpW measurements (3-5nm). The majority of traces fall in to the second cluster (n=21). Here traces are detected in which the BBL unfolds via several steps within an overall sigmoidal shape unfolding signal (panel B). In the third cluster (n=4), the BBL unfolds via a single step (panel C). However also this single step consists of multiple noisy steps (see insert in panal C) and therefore is not a clear unfolding step like detected for the Csp or the titin I27 domain. The BBL domain does not show a clear repeatable unfolding pattern and is quite comparable with the mechanical description observed within the corresponding measurements of the gpW protein.

List of publications

Manuscripts under preparation:

- Single-Molecule Analysis of a Two-State Folding Protein Unveils a Stochastic Multi-step Mechanical Unfolding Process that is Tunable by Force.Jörg Schönfelder, Raúl Pérez Jiménez and Victor Muñoz (Submitted to PNAS)

Bibliography

[1] C.B. Anfinsen. Principles that Govern the Folding of Protein Chains. *Science*, 181(4096):223–230, 1973.

[2] Edgar Haber and Christian B Anfinsen. Side-chain Interactions Governing the Pairing of Half-cystine Residues in Ribonuclease Side-chain Interactions Governing the Pairing Residues in Ribonuclease *. 1962.

[3] A I Bartlett and S E Radford. An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nat Struct Mol Biol*, 16(6):582–588, 2009.

[4] Ken A Dill, Justin L Maccallum, and Protein Folding. The Protein-Folding Problem , 50 Years On. *Science*, pages 1042–1047, 2012.

[5] Ken a Dill, S Banu Ozkan, M Scott Shell, and Thomas R Weikl. The protein folding problem. *Annual review of biophysics*, 37:289–316, January 2008.

[6] Robert Zwanzig, Attila Szabo, and Biman Bagchi. Levinthal ' s paradox. 89(January):20–22, 1992.

[7] T Y Tsong, R L Baldwin, and E L Elson. The sequential unfolding of ribonuclease A: detection of a fast initial phase in the kinetics of unfolding. *Proceedings of the National Academy of Sciences of the United States of America*, 68(11):2712–5, November 1971.

[8] ATSUSHI IKAI and CHARLES TANFORD. Kinetic Evidence for Incorrectly Folded Intermediate States in the Refolding of Denatured Proteins. *Nature*, 230(5289):100–102, March 1971.

[9] H Roder, G A Elöve, and S W Englander. Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. *Nature*, 335(6192):700–4, October 1988.

[10] A D Miranker and C M Dobson. Collapse and cooperativity in protein folding. *Current opinion in structural biology*, 6(1):31–42, February 1996.

[11] Robert Miller, Craig A. Danko, Michael J. Fasolka, Anna C. Balazs, Hue Sun Chan, and Ken A. Dill. Folding kinetics of proteins and copolymers. *The Journal of Chemical Physics*, 96(1):768, January 1992.

[12] H S Chan, S Bromberg, and K A Dill. Models of cooperativity in protein folding. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 348(1323):61–70, April 1995.

[13] J D Bryngelson, J N Onuchic, N D Socci, and P G Wolynes. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins*, 21(3):167–95, March 1995.

[14] K A Dill and H S Chan. From Levinthal to pathways to funnels. *Nature structural biology*, 4(1):10–9, January 1997.

[15] José Nelson Onuchic, Zaida Luthey-schulten, and Peter G Wolynes. THEORY OF PROTEIN FOLDING : The Energy Landscape Perspective. (1):545–600, 1997.

[16] J. D. Bryngelson and P. G. Wolynes. Spin glasses and the statistical mechanics of protein folding. *Proceedings of the National Academy of Sciences*, 84(21):7524–7528, November 1987.

[17] José Nelson Onuchic and Peter G Wolynes. Theory of protein folding. *Current opinion in structural biology*, 14(1):70–5, February 2004.

[18] S Arrhenius. On the reaction rate of the inversion of non-refined sugar upon souring. *Z Phys Chem*, 1889.

[19] Henry Eyring. The Activated Complex in Chemical Reactions. *The Journal of Chemical Physics*, 3(2):107, November 1935.

[20] H.A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, April 1940.

[21] Peter Hänggi and Michal Borkovec. Reaction-rate theory: fifty years after Kramers. *Reviews of Modern Physics*, 62(2):251–341, April 1990.

[22] Victor Muñoz. Thermodynamics and kinetics of downhill protein folding investigated with a simple statistical mechanical model. *International Journal of Quantum Chemistry*, 90(4-5):1522–1528, October 2002.

[23] K W Plaxco, K T Simons, and D Baker. Contact order, transition state placement and the refolding rates of single domain proteins. *Journal of molecular biology*, 277(4):985–94, April 1998.

[24] A R Dinner, A Sali, L J Smith, C M Dobson, and M Karplus. Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem Sci*, 25(7):331–339, 2000.

[25] Victor Muñoz, Mourad Sadqi, Athi N Naganathan, and David de Sancho. Exploiting the downhill folding regime via experiment. *HFSP journal*, 2(6):342–53, December 2008.

[26] N Go. Theoretical studies of protein folding. *Annual review of biophysics and bioengineering*, 12:183–210, January 1983.

[27] Robert B Best. Atomistic molecular simulations of protein folding. *Current opinion in structural biology*, 22(1):52–61, February 2012.

[28] Michael Levitt and Arieh Warshel. Computer simulation of protein folding. *Nature*, 253(5494):694–698, February 1975.

[29] J A McCammon, B R Gelin, and M Karplus. Dynamics of folded proteins. *Nature*, 267(5612):585–90, June 1977.

[30] Alexander D Mackerell. Empirical force fields for biological macromolecules: overview and issues. *Journal of computational chemistry*, 25(13):1584–604, October 2004.

[31] David E. Shaw, Kevin J. Bowers, Edmond Chow, Michael P. Eastwood, Douglas J. Ierardi, John L. Klepeis, Jeffrey S. Kuskin, Richard H. Larson, Kresten Lindorff-Larsen, Paul Maragakis, Mark A. Moraes, Ron O. Dror, Stefano Piana, Yibing Shan, Brian Towles, John K. Salmon, J. P. Grossman, Kenneth M. Mackenzie, Joseph A. Bank, Cliff Young, Martin M. Deneroff, and Brannon Batson. Millisecond-scale molecular dynamics simulations on Anton. In *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis - SC '09*, page 1, New York, New York, USA, 2009. ACM Press.

[32] David E Shaw, Paul Maragakis, Kresten Lindorff-Larsen, Stefano Piana, Ron O Dror, Michael P Eastwood, Joseph A Bank, John M Jumper, John K Salmon, Yibing Shan, and Willy Wriggers. Atomic-level characterization of the structural dynamics of proteins. *Science (New York, N.Y.)*, 330(6002):341–6, October 2010.

[33] Kresten Lindorff-Larsen, Stefano Piana, Ron O Dror, and David E Shaw. How fast-folding proteins fold. *Science (New York, N.Y.)*, 334(6055):517–20, October 2011.

[34] Eric R Henry, Robert B Best, and William A Eaton. Comparing a simple theoretical model for protein folding with all-atom molecular dynamics simulations. *Proceedings of the National Academy of Sciences of the United States of America*, 110(44):17880–5, October 2013.

[35] Lorenzo Sborgi, Abhinav Verma, Stefano Piana, Kresten Lindorff-Larsen, Michele Cerminara, Clara M Santiveri, David E Shaw, Eva de Alba, and Victor Muñoz. Interaction Networks in Protein Folding via Atomic-Resolution Experiments and Long-Time-Scale Molecular Dynamics Simulations. *Journal of the American Chemical Society*, May 2015.

[36] Thomas J Lane, Diwakar Shukla, Kyle A Beauchamp, and Vijay S Pande. To milliseconds and beyond: challenges in the simulation of protein folding. *Current opinion in structural biology*, 23(1):58–65, February 2013.

[37] Mark S Friedrichs, Peter Eastman, Vishal Vaidyanathan, Mike Houston, Scott Legrand, Adam L Beberg, Daniel L Ensign, Christopher M Bruns, and Vijay S Pande. Accelerating molecular dynamic simulation on graphics processing units. *Journal of computational chemistry*, 30(6):864–72, April 2009.

[38] Hai Nguyen, James Maier, He Huang, Victoria Perrone, and Carlos Simmerling. Folding simulations for proteins with diverse topologies are accessible in days with a physics-based force field and implicit solvent. *Journal of the American Chemical Society*, 136(40):13959–62, October 2014.

[39] S E Jackson and A R Fersht. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry*, 30(43):10428–35, October 1991.

[40] Victor Muñoz and Jose M Sanchez-Ruiz. Exploring protein-folding ensembles: a variable-barrier model for the analysis of equilibrium unfolding experiments. *Pro-*

ceedings of the National Academy of Sciences of the United States of America, 101(51):17646–51, December 2004.

[41] Athi N Naganathan, Jose M Sanchez-Ruiz, and Victor Muñoz. Direct measurement of barrier heights in protein folding. *Journal of the American Chemical Society*, 127(51):17970–1, December 2005.

[42] R Godoy-Ruiz, E R Henry, J Kubelka, J Hofrichter, V Munoz, J M Sanchez-Ruiz, and W A Eaton. Estimating free-energy barrier heights for an ultrafast folding protein from calorimetric and kinetic data. *J Phys Chem B*, 112(19):5938–5949, 2008.

[43] R Callender. Probing protein dynamics using temperature jump relaxation spectroscopy. *Current Opinion in Structural Biology*, 12(5):628–633, October 2002.

[44] S Rozovsky and A E McDermott. The time scale of the catalytic loop motion in triosephosphate isomerase. *Journal of molecular biology*, 310(1):259–70, June 2001.

[45] S Rozovsky, G Jogl, L Tong, and A E McDermott. Solution-state NMR investigations of triosephosphate isomerase active site loop motion: ligand release in relation to active site loop dynamics. *Journal of molecular biology*, 310(1):271–80, June 2001.

[46] Peggy A. Thompson, Victor Muñoz, Gouri S. Jas, Eric R. Henry, William A. Eaton, and James Hofrichter. The Helix-Coil Kinetics of a Heteropeptide. *The Journal of Physical Chemistry B*, 104(2):378–389, January 2000.

[47] S Williams, T P Causgrove, R Gilmanshin, K S Fang, R H Callender, W H Woodruff, and R B Dyer. Fast events in protein folding: helix melting and formation in a small peptide. *Biochemistry*, 35(3):691–7, January 1996.

[48] V Muñoz, P A Thompson, J Hofrichter, and W A Eaton. Folding dynamics and mechanism of beta-hairpin formation. *Nature*, 390(6656):196–9, November 1997.

[49] Mourad Sadqi, Lisa J Lapidus, and Victor Mun. How fast is protein hydrophobic collapse ? 100(21), 2003.

[50] Victor Muñoz. Conformational dynamics and ensembles in protein folding. *Annual review of biophysics and biomolecular structure*, 36:395–412, January 2007.

[51] Jan Kubelka, James Hofrichter, and William a Eaton. The protein folding 'speed limit'. *Current opinion in structural biology*, 14(1):76–88, February 2004.

[52] R Brian Dyer. Ultrafast and downhill protein folding. *Current opinion in structural biology*, 17(1):38–47, February 2007.

[53] Jan Kubelka, Thang K Chiu, David R Davies, William A Eaton, and James Hofrichter. Sub-microsecond protein folding. *Journal of molecular biology*, 359(3):546–53, June 2006.

[54] Jeffrey K Myers and Terrence G Oas. Mechanism of fast protein folding. *Annual review of biochemistry*, 71:783–815, January 2002.

[55] Wei Yuan Yang and Martin Gruebele. Folding at the speed limit. *Nature*, 423(6936):193–7, May 2003.

[56] David De Sancho, Urmi Doshi, and Victor Muñoz. Protein folding rates and stability: how much is there beyond size? *Journal of the American Chemical Society*, 131(6):2074–5, February 2009.

[57] David De Sancho and Victor Muñoz. Integrated prediction of protein folding and unfolding rates from only size and structural class. *Physical chemistry chemical physics : PCCP*, 13(38):17030–43, October 2011.

[58] Athi N Naganathan, Urmi Doshi, and Victor Muñoz. Protein folding kinetics: barrier effects in chemical and thermal denaturation experiments. *Journal of the American Chemical Society*, 129(17):5673–82, May 2007.

[59] V Muñoz and W A Eaton. A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proceedings of the National Academy of Sciences of the United States of America*, 96(20):11311–6, September 1999.

[60] Victor Muñoz. A simple theoretical model goes a long way in explaining complex behavior in protein folding. *Proceedings of the National Academy of Sciences of the United States of America*, 111(45):15863–4, November 2014.

[61] A C Ferreon and A A Deniz. Protein folding at single-molecule resolution. *Biochim Biophys Acta*, 1814(8):1021–1029, 2011.

[62] Alessandro Borgia, Philip M Williams, and Jane Clarke. Single-molecule studies of protein folding. *Annual review of biochemistry*, 77:101–25, January 2008.

[63] Benjamin Schuler. Single-molecule FRET of protein structure and dynamics - a primer. *Journal of nanobiotechnology*, 11 Suppl 1:S2, January 2013.

[64] B Schuler and W A Eaton. Protein folding studied by single-molecule FRET. *Curr Opin Struct Biol*, 18(1):16–26, 2008.

[65] Keir C Neuman and Attila Nagy. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nature methods*, 5(6):491–505, June 2008.

[66] Rahul Roy, Sungchul Hohng, and Taekjip Ha. A practical guide to single-molecule FRET. *Nature methods*, 5(6):507–16, June 2008.

[67] E Rhoades, M Cohen, B Schuler, and G Haran. Two-state folding observed in individual protein molecules. *J Am Chem Soc*, 126(45):14686–14687, 2004.

[68] Hoi Sung Chung, Kevin McHale, John M Louis, and William A Eaton. Single-molecule fluorescence experiments determine protein folding transition path times. *Science (New York, N.Y.)*, 335(6071):981–4, February 2012.

[69] Hoi Sung Chung and William A Eaton. Single-molecule fluorescence probes dynamics of barrier crossing. *Nature*, 502(7473):685–8, October 2013.

[70] Evangelos Sisamakis, Alessandro Valeri, Stanislav Kalinin, Paul J Rothwell, and Claus A M Seidel. Accurate single-molecule FRET studies using multiparameter fluorescence detection. *Methods in enzymology*, 475:455–514, January 2010.

[71] S E Jackson. How do small single-domain proteins fold? *Folding & design*, 3(4):R81–91, January 1998.

[72] Karen L Maxwell, David Wildes, Arash Zarrine-Afsar, Miguel A De Los Rios, Andrew G Brown, Claire T Friel, Linda Hedberg, Jia-Cherng Horng, Diane Bona, Erik J Miller, Alexis Vallée-Bélisle, Ewan R G Main, Francesco Bemporad, Linlin Qiu, Kaare Teilum, Ngoc-Diep Vu, Aled M Edwards, Ingo Ruczinski, Flemming M Poulsen, Birthe B Kragelund, Stephen W Michnick, Fabrizio Chiti, Yawen Bai, Stephen J Hagen, Luis Serrano, Mikael Oliveberg, Daniel P Raleigh, Pernilla Wittung-Stafshede, Sheena E Radford, Sophie E Jackson, Tobin R Sosnick, Susan Marqusee, Alan R Davidson, and Kevin W Plaxco. Protein folding: defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein science : a publication of the Protein Society*, 14(3):602–16, March 2005.

[73] B Schuler, E A Lipman, and W A Eaton. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature*, 419(6908):743–747, 2002.

[74] Luis A Campos, Mourad Sadqi, Jianwei Liu, Xiang Wang, Douglas S English, and Victor Muñoz. Gradual disordering of the native state on a slow two-state folding protein monitored by single-molecule fluorescence spectroscopy and NMR. *The journal of physical chemistry. B*, 117(42):13120–31, October 2013.

[75] W. A. Eaton. Searching for "downhill scenarios" in protein folding. *Proceedings of the National Academy of Sciences*, 96(11):5897–5899, May 1999.

[76] M M Garcia-Mira, M Sadqi, N Fischer, J M Sanchez-Ruiz, and V Munoz. Experimental identification of downhill protein folding. *Science*, 298(5601):2191–2195, 2002.

[77] Peng Li, Fabiana Y Oliva, Athi N Naganathan, and Victor Mun. Dynamics of one-state downhill protein folding. 106(1), 2009.

[78] Adam Fung, Peng Li, Raquel Godoy-Ruiz, Jose M Sanchez-Ruiz, and Victor Muñoz. Expanding the realm of ultrafast protein folding: gpW, a midsize natural single-domain with alpha+beta topology that folds downhill. *Journal of the American Chemical Society*, 130(23):7489–95, June 2008.

[79] Celia Sanchez-Medina, Ashok Sekhar, Pramodh Vallurupalli, Michele Cerminara, Victor Muñoz, and Lewis E Kay. Probing the free energy landscape of the fast-folding gpW protein by relaxation dispersion NMR. *Journal of the American Chemical Society*, 136(20):7444–51, May 2014.

[80] Hairong Ma and Martin Gruebele. Kinetics are probe-dependent during downhill folding of an engineered lambda6-85 protein. *Proceedings of the National Academy of Sciences of the United States of America*, 102(7):2283–7, February 2005.

[81] D T Leeson, F Gai, H M Rodriguez, L M Gregoret, and R B Dyer. Protein folding and unfolding on a complex energy landscape. *Proc Natl Acad Sci U S A*, 97(6):2527–2532, 2000.

[82] Santosh Kumar Jha, Deepak Dhar, Guruswamy Krishnamoorthy, and Jayant B Udgaonkar. Continuous dissolution of structure during the unfolding of a small

protein. *Proceedings of the National Academy of Sciences of the United States of America*, 106(27):11113–8, July 2009.

[83] Mourad Sadqi, David Fushman, and Victor Muñoz. Atom-by-atom analysis of global downhill protein folding. *Nature*, 442(7100):317–21, July 2006.

[84] Jianwei Liu, Luis A Campos, Michele Cerminara, Xiang Wang, Ravishankar Ramanathan, Douglas S English, and Victor Muñoz. Exploring one-state downhill protein folding in single molecules. *Proceedings of the National Academy of Sciences of the United States of America*, 109(1):179–84, January 2012.

[85] Luis A Campos, Jianwei Liu, Xiang Wang, Ravishankar Ramanathan, Douglas S English, and Victor Muñoz. A photoprotection strategy for microsecond-resolution single-molecule fluorescence spectroscopy. *Nature methods*, 8(2):143–6, February 2011.

[86] Carlos Bustamante, Yann R Chemla, Nancy R Forde, and David Izhaky. Mechanical processes in biochemistry. *Annual review of biochemistry*, 73:705–48, January 2004.

[87] Wolfgang A Linke and Anika Grützner. Pulling single molecules of titin by AFM–recent advances and physiological implications. *Pflügers Archiv : European journal of physiology*, 456(1):101–15, April 2008.

[88] Tania A Baker and Robert T Sauer. ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochimica et biophysica acta*, 1823(1):15–28, January 2012.

[89] A Ashkin, J M Dziedzic, J E Bjorkholm, and S Chu. Observation of a single-beam gradient force optical trap for dielectric particles. *Optics letters*, 11(5):288, May 1986.

[90] K Svoboda and S M Block. Biological applications of optical forces. *Annual review of biophysics and biomolecular structure*, 23:247–85, January 1994.

[91] Jeffrey R Moffitt, Yann R Chemla, Steven B Smith, and Carlos Bustamante. Recent advances in optical tweezers. *Annual review of biochemistry*, 77:205–28, January 2008.

[92] Matthew J Lang, Charles L Asbury, Joshua W Shaevitz, and Steven M Block. An automated two-dimensional optical force clamp for single molecule studies. *Biophysical journal*, 83(1):491–501, July 2002.

[93] Gabriel Žoldák, Johannes Stigler, Benjamin Pelz, Hongbin Li, and Matthias Rief. Ultrafast folding kinetics and cooperativity of villin headpiece in single-molecule force spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 110(45):18156–61, November 2013.

[94] Charlie Gosse and Vincent Croquette. Magnetic tweezers: micromanipulation and force measurement at the molecular level. *Biophysical journal*, 82(6):3314–29, June 2002.

[95] G. Binnig and C. F. Quate. Atomic Force Microscope. *Physical Review Letters*, 56(9):930–933, March 1986.

[96] G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel. Surface Studies by Scanning Tunneling Microscopy. *Physical Review Letters*, 49(1):57–61, July 1982.

[97] Z Shao, J Yang, and A P Somlyo. Biological atomic force microscopy: from microns to nanometers and beyond. *Annual review of cell and developmental biology*, 11:241–65, January 1995.

[98] Carlos Bustamante, Claudio Rivetti, and David J Keller. Scanning force microscopy under aqueous solutions. *Current Opinion in Structural Biology*, 7(5):709–716, October 1997.

[99] Thomas F Keller, Jörg Schönfelder, Jörg Reichert, Nunzio Tuccitto, Antonino Licciardello, Grazia M L Messina, Giovanni Marletta, and Klaus D Jandt. How the surface nanostructure of polyethylene affects protein assembly and orientation. *ACS nano*, 5(4):3120–31, April 2011.

[100] Iwan A T Schaap, Carolina Carrasco, Pedro J de Pablo, and Christoph F Schmidt. Kinesin walks the line: single motors observed by atomic force microscopy. *Biophysical journal*, 100(10):2450–6, May 2011.

[101] Livan Alonso-Sarduy, Giovanni Longo, Giovanni Dietler, and Sandor Kasas. Time-lapse AFM imaging of DNA conformational changes induced by daunorubicin. *Nano letters*, 13(11):5679–84, January 2013.

[102] Noriyuki Kodera, Daisuke Yamamoto, Ryoki Ishikawa, and Toshio Ando. Video imaging of walking myosin V by high-speed atomic force microscopy. *Nature*, 468(7320):72–6, November 2010.

[103] Hayato Yamashita, Azuma Taoka, Takayuki Uchihashi, Tomoya Asano, Toshio Ando, and Yoshihiro Fukumori. Single-molecule imaging on living bacterial cell surface by high-speed AFM. *Journal of molecular biology*, 422(2):300–9, September 2012.

[104] G U Lee, L A Chrisey, and R J Colton. Direct measurement of the forces between complementary strands of DNA. *Science (New York, N.Y.)*, 266(5186):771–3, November 1994.

[105] M Rief, M Gautel, F Oesterhelt, J M Fernandez, and H E Gaub. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science (New York, N.Y.)*, 276(5315):1109–12, May 1997.

[106] H Delanoë-Ayari, R Al Kurdi, M Vallade, D Gulino-Debrac, and D Riveline. Membrane and acto-myosin tension promote clustering of adhesion proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 101(8):2229–34, February 2004.

[107] Peter Hinterdorfer and Yves F Dufrêne. Detection and localization of single molecular recognition events using atomic force microscopy. *Nature methods*, 3(5):347–55, May 2006.

[108] Ionel Popa, Ronen Berkovich, Jorge Alegre-Cebollada, Carmen L Badilla, Jaime Andrés Rivas-Pardo, Yukinori Taniguchi, Masaru Kawakami, and Julio M Fernandez. Nanomechanics of HaloTag tethers. *Journal of the American Chemical Society*, 135(34):12762–71, August 2013.

[109] Jeffrey L. Hutter and John Bechhoefer. Calibration of atomic-force microscope tips. *Review of Scientific Instruments*, 64(7):1868, July 1993.

[110] O. Kratky and G. Porod. Röntgenuntersuchung gelöster Fadenmoleküle. *Recueil des Travaux Chimiques des Pays-Bas*, 68(12):1106–1122, September 1949.

[111] C Bustamante, J F Marko, E D Siggia, and S Smith. Entropic elasticity of lambda-phage DNA. *Science (New York, N.Y.)*, 265(5178):1599–600, September 1994.

[112] John F. Marko and Eric D. Siggia. Stretching DNA. *Macromolecules*, 28(26):8759–8770, December 1995.

[113] Werner Kuhn and F. Grün. Beziehungen zwischen elastischen Konstanten und Dehnungsdoppelbrechung hochelastischer Stoffe. *Kolloid-Zeitschrift*, 101(3):248–271, December 1942.

[114] A F Oberhauser, P K Hansma, M Carrion-Vazquez, and J M Fernandez. Stepwise unfolding of titin under force-clamp atomic force microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2):468–72, January 2001.

[115] I Popa, P Kosuri, J Alegre-Cebollada, S Garcia-Manyes, and J M Fernandez. Force dependency of biochemical reactions measured by single-molecule force-clamp spectroscopy. *Nat Protoc*, 8(7):1261–1276, 2013.

[116] M Schlierf, H Li, and J M Fernandez. The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. *Proc Natl Acad Sci U S A*, 101(19):7299–7304, 2004.

[117] Albert Galera-Prat, Angel Gómez-Sicilia, Andres F Oberhauser, Marek Cieplak, and Mariano Carrión-Vázquez. Understanding biology by stretching proteins: recent progress. *Current opinion in structural biology*, 20(1):63–9, March 2010.

[118] Gwangrog Lee, Khadar Abdi, Yong Jiang, Peter Michaely, Vann Bennett, and Piotr E Marszalek. Nanospring behaviour of ankyrin repeats. *Nature*, 440(7081):246–9, March 2006.

[119] M Rief, J Pascual, M Saraste, and H E Gaub. Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles. *Journal of molecular biology*, 286(2):553–61, February 1999.

[120] Alejandro Valbuena, Andrés Manuel Vera, Javier Oroz, Margarita Menéndez, and Mariano Carrión-Vázquez. Mechanical properties of β-catenin revealed by single-molecule experiments. *Biophysical journal*, 103(8):1744–52, October 2012.

[121] Armando del Rio, Raul Perez-Jimenez, Ruchuan Liu, Pere Roca-Cusachs, Julio M Fernandez, and Michael P Sheetz. Stretching single talin rod molecules activates vinculin binding. *Science (New York, N.Y.)*, 323(5914):638–41, January 2009.

[122] M. Rief. Reversible Unfolding of Individual Titin Immunoglobulin Domains by AFM. *Science*, 276(5315):1109–1112, May 1997.

[123] Alejandro Valbuena, Javier Oroz, Rubén Hervás, Andrés Manuel Vera, David Rodríguez, Margarita Menéndez, Joanna I Sulkowska, Marek Cieplak, and Mariano Carrión-Vázquez. On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. *Proceedings of the National Academy of Sciences of the United States of America*, 106(33):13791–6, August 2009.

[124] Hendrik Dietz and Matthias Rief. Exploring the energy landscape of GFP by single-molecule mechanical experiments. *Proceedings of the National Academy of Sciences of the United States of America*, 101(46):16192–7, November 2004.

[125] Jorge Alegre-Cebollada, Carmen L Badilla, and Julio M Fernández. Isopeptide bonds block the mechanical extension of pili in pathogenic Streptococcus pyogenes. *The Journal of biological chemistry*, 285(15):11235–42, April 2010.

[126] A F Oberhauser, P E Marszalek, H P Erickson, and J M Fernandez. The molecular elasticity of the extracellular matrix protein tenascin. *Nature*, 393(6681):181–5, May 1998.

[127] M Grandbois, M Beyer, M Rief, H Clausen-Schaumann, and HE Gaub. How strong is a covalent bond? *Science (New York, N.Y.)*, 283(5408):1727–30, March 1999.

[128] T Hoffmann, K M Tych, D J Brockwell, and L Dougan. Single-molecule force spectroscopy identifies a small cold shock protein as being mechanically robust. *J Phys Chem B*, 117(6):1819–1826, 2013.

[129] David J Brockwell, Godfrey S Beddard, Emanuele Paci, Dan K West, Peter D Olmsted, D Alastair Smith, and Sheena E Radford. Mechanically unfolding the small, topologically simple protein L. *Biophysical journal*, 89(1):506–19, July 2005.

[130] Mateusz Sikora, Joanna I Sulkowska, Bartlomiej S Witkowski, and Marek Cieplak. BSDB: the biomolecule stretching database. *Nucleic acids research*, 39(Database issue):D443–50, January 2011.

[131] M Carrion-Vazquez, A F Oberhauser, T E Fisher, P E Marszalek, H Li, and J M Fernandez. Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. *Prog Biophys Mol Biol*, 74(1-2):63–91, 2000.

[132] L Li, H H Huang, C L Badilla, and J M Fernandez. Mechanical unfolding intermediates observed by single-molecule force spectroscopy in a fibronectin type III module. *J Mol Biol*, 345(4):817–826, 2005.

[133] Q Peng and H Li. Atomic force microscopy reveals parallel mechanical unfolding pathways of T4 lysozyme: evidence for a kinetic partitioning mechanism. *Proc Natl Acad Sci U S A*, 105(6):1885–1890, 2008.

[134] P E Marszalek, H Lu, H Li, M Carrion-Vazquez, A F Oberhauser, K Schulten, and J M Fernandez. Mechanical unfolding intermediates in titin modules. *Nature*, 402(6757):100–103, 1999.

[135] Elio A Abbondanzieri, William J Greenleaf, Joshua W Shaevitz, Robert Landick, and Steven M Block. Direct observation of base-pair stepping by RNA polymerase. *Nature*, 438(7067):460–5, November 2005.

[136] William J Greenleaf, Kirsten L Frieda, Daniel A N Foster, Michael T Woodside, and Steven M Block. Direct observation of hierarchical folding in single riboswitch aptamers. *Science (New York, N.Y.)*, 319(5863):630–3, February 2008.

[137] Marie-Eve Aubin-Tam, Adrian O Olivares, Robert T Sauer, Tania A Baker, and Matthew J Lang. Single-molecule protein unfolding and translocation by an ATP-fueled proteolytic machine. *Cell*, 145(2):257–67, April 2011.

[138] José A Morin, Francisco J Cao, José M Lázaro, J Ricardo Arias-Gonzalez, José M Valpuesta, José L Carrascosa, Margarita Salas, and Borja Ibarra. Active DNA unwinding dynamics during processive DNA replication. *Proceedings of the National Academy of Sciences of the United States of America*, 109(21):8115–20, May 2012.

[139] Carolina Carrasco, Neville S Gilhooly, Mark S Dillingham, and Fernando Moreno-Herrero. On the mechanism of recombination hotspot scanning during double-stranded DNA break resection. *Proceedings of the National Academy of Sciences of the United States of America*, 110(28):E2562–71, July 2013.

[140] M S Kellermayer, S B Smith, H L Granzier, and C Bustamante. Folding-unfolding transitions in single titin molecules characterized with laser tweezers. *Science (New York, N.Y.)*, 276(5315):1112–6, May 1997.

[141] C Cecconi, E A Shank, C Bustamante, and S Marqusee. Direct observation of the three-state folding of a single protein molecule. *Science*, 309(5743):2057–2060, 2005.

[142] J Christof M Gebhardt, Thomas Bornschlögl, and Matthias Rief. Full distance-resolved folding energy landscape of one single protein molecule. *Proceedings of*

the National Academy of Sciences of the United States of America, 107(5):2013–8, February 2010.

[143] Hao Yu, Xia Liu, Krishna Neupane, Amar Nath Gupta, Angela M Brigley, Allison Solanki, Iveta Sosova, and Michael T Woodside. Direct observation of multiple misfolding pathways in a single prion protein molecule. *Proceedings of the National Academy of Sciences of the United States of America*, 109(14):5283–8, April 2012.

[144] Ying Gao, Sylvain Zorman, Gregory Gundersen, Zhiqun Xi, Lu Ma, George Sirinakis, James E Rothman, and Yongli Zhang. T (where k. 28, 2012.

[145] Bharat Jagannathan, Phillip J Elms, Carlos Bustamante, and Susan Marqusee. Direct observation of a force-induced switch in the anisotropic mechanical unfolding pathway of a protein. *Proceedings of the National Academy of Sciences of the United States of America*, 109(44):17820–5, October 2012.

[146] B Jagannathan and S Marqusee. Protein folding and unfolding under force. *Biopolymers*, 99(11):860–869, 2013.

[147] Julio M Fernandez and Hongbin Li. Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science (New York, N.Y.)*, 303(5664):1674–8, March 2004.

[148] Jan Philipp Junker, Fabian Ziegler, and Matthias Rief. Ligand-dependent equilibrium fluctuations of single calmodulin molecules. *Science (New York, N.Y.)*, 323(5914):633–7, January 2009.

[149] Jan Philipp Junker and Matthias Rief. Single-molecule force spectroscopy distinguishes target binding modes of calmodulin. *Proceedings of the National Academy of Sciences of the United States of America*, 106(34):14361–6, August 2009.

[150] Sergi Garcia-Manyes, Jasna Brujić, Carmen L Badilla, and Julio M Fernández. Force-clamp spectroscopy of single-protein monomers reveals the individual unfolding and folding pathways of I27 and ubiquitin. *Biophysical journal*, 93(7):2436–46, October 2007.

[151] S Garcia-Manyes, L Dougan, C L Badilla, J Brujic, and J M Fernandez. Direct observation of an ensemble of stable collapsed states in the mechanical folding of ubiquitin. *Proc Natl Acad Sci U S A*, 106(26):10534–10539, 2009.

[152] Gabriel Zoldák and Matthias Rief. Force as a single molecule probe of multidimensional protein energy landscapes. *Current opinion in structural biology*, 23(1):48–57, February 2013.

[153] Ignacio Tinoco and Carlos Bustamante. The effect of force on thermodynamics and kinetics of single molecule reactions. *Biophysical chemistry*, 101-102:513–33, December 2002.

[154] G I Bell. Models for the specific adhesion of cells to cells. *Science*, 200(4342):618–627, 1978.

[155] M T Woodside and S M Block. Reconstructing folding energy landscapes by single-molecule force spectroscopy. *Annu Rev Biophys*, 43:19–39, 2014.

[156] Michael T Woodside, Peter C Anthony, William M Behnke-Parks, Kevan Larizadeh, Daniel Herschlag, and Steven M Block. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science (New York, N.Y.)*, 314(5801):1001–4, November 2006.

[157] Tzu-Ling Kuo, Sergi Garcia-Manyes, Jingyuan Li, Itay Barel, Hui Lu, Bruce J Berne, Michael Urbakh, Joseph Klafter, and Julio M Fernández. Probing static disorder in Arrhenius kinetics by single-molecule force spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 107(25):11336–40, June 2010.

[158] Johannes Stigler, Fabian Ziegler, Anja Gieseke, J Christof M Gebhardt, and Matthias Rief. The complex folding network of single calmodulin molecules. *Science (New York, N.Y.)*, 334(6055):512–6, October 2011.

[159] Ronen Berkovich, Sergi Garcia-Manyes, Michael Urbakh, Joseph Klafter, and Julio M Fernandez. Collapse dynamics of single proteins extended by force. *Biophysical journal*, 98(11):2692–701, June 2010.

[160] Ronen Berkovich, Sergi Garcia-Manyes, Joseph Klafter, Michael Urbakh, and Julio M Fernández. Hopping around an entropic barrier created by force. *Biochemical and biophysical research communications*, 403(1):133–7, December 2010.

[161] Olga K Dudko, Thomas G W Graham, and Robert B Best. Locating the barrier for folding of single molecules under an external force. *Physical review letters*, 107(20):208301, November 2011.

[162] G Z Genchev, M Kallberg, G Gursoy, A Mittal, L Dubey, O Perisic, G Feng, R Langlois, and H Lu. Mechanical signaling on the single protein level studied using steered molecular dynamics. *Cell Biochem Biophys*, 55(3):141–152, 2009.

[163] Eric H Lee, Jen Hsin, Marcos Sotomayor, Gemma Comellas, and Klaus Schulten. Discovery through the computational microscope. *Structure (London, England : 1993)*, 17(10):1295–306, October 2009.

[164] Mariano Carrion-Vazquez, Hongbin Li, Hui Lu, Piotr E Marszalek, Andres F Oberhauser, and Julio M Fernandez. The mechanical stability of ubiquitin is linkage dependent. *Nature structural biology*, 10(9):738–43, September 2003.

[165] Jen Hsin, Johan Strümpfer, Eric H Lee, and Klaus Schulten. Molecular origin of the hierarchical elasticity of titin: simulation, experiment, and theory. *Annual review of biophysics*, 40:187–203, January 2011.

[166] F Rico, L Gonzalez, I Casuso, M Puig-Vidal, and S Scheuring. High-speed force spectroscopy unfolds titin at the velocity of molecular dynamics simulations. *Science*, 342(6159):741–743, 2013.

[167] M. Carrion-Vazquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke, and J. M. Fernandez. Mechanical and chemical unfolding of a single protein: A comparison. *Proceedings of the National Academy of Sciences*, 96(7):3694–3699, March 1999.

[168] S Improta, A S Politou, and A Pastore. Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. *Structure (London, England : 1993)*, 4(3):323–37, March 1996.

[169] A S Politou, D J Thomas, and A Pastore. The folding and stability of titin immunoglobulin-like modules, with implications for the mechanism of elasticity. *Biophysical journal*, 69(6):2601–10, December 1995.

[170] D Giganti, J Alegre-Cebollada, S Urresti, D Albesa-Jove, A Rodrigo-Unzueta, N Comino, M Kachala, S Lopez-Fernandez, D I Svergun, J M Fernandez, and M E Guerin. Conformational plasticity of the essential membrane-associated mannosyltransferase PimA from mycobacteria. *J Biol Chem*, 288(41):29797–29808, 2013.

[171] Hu Chen, Guohua Yuan, Ricksen Winardhi, Mingxi Yao, Ionel Popa, Julio M Fernandez, and Jie Yan. Dynamics of equilibrium folding and unfolding transitions

of titin immunoglobulin domain under constant forces. *Journal of the American Chemical Society*, 137(10):3540–3546, March 2015.

[172] S Khorasanizadeh, I D Peters, and H Roder. Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nature structural biology*, 3(2):193–205, February 1996.

[173] C M Pickart. Mechanisms underlying ubiquitination. *Annual review of biochemistry*, 70:503–33, January 2001.

[174] J Oroz, R Hervas, and M Carrion-Vazquez. Unequivocal single-molecule force spectroscopy of proteins by AFM using pFS vectors. *Biophys J*, 102(3):682–690, 2012.

[175] M A Robien, G M Clore, J G Omichinski, R N Perham, E Appella, K Sakaguchi, and A M Gronenborn. Three-dimensional solution structure of the E3-binding domain of the dihydrolipoamide succinyltransferase core from the 2-oxoglutarate dehydrogenase multienzyme complex of Escherichia coli. *Biochemistry*, 31(13):3463–71, April 1992.

[176] Sowmini Kumaran, Mulchand S Patel, and Frank Jordan. Nuclear magnetic resonance approaches in the study of 2-oxo acid dehydrogenase multienzyme complexes–a literature review. *Molecules (Basel, Switzerland)*, 18(10):11873–903, January 2013.

[177] R N Perham. Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. *Annual review of biochemistry*, 69:961–1004, January 2000.

[178] Tanay M Desai, Michele Cerminara, Mourad Sadqi, and Victor Muñoz. The effect of electrostatics on the marginal cooperativity of an ultrafast folding protein. *The Journal of biological chemistry*, 285(45):34549–56, November 2010.

[179] Michele Cerminara, Tanay M Desai, Mourad Sadqi, and Victor Muñoz. Downhill protein folding modules as scaffolds for broad-range ultrafast biosensors. *Journal of the American Chemical Society*, 134(19):8010–3, May 2012.

[180] Michele Cerminara, Luis A Campos, Ravishankar Ramanathan, and Victor Muñoz. Slow proton transfer coupled to unfolding explains the puzzling results of single-molecule experiments on BBL, a paradigmatic downhill folding protein. *PloS one*, 8(10):e78044, January 2013.

[181] Samuel S Cho, Patrick Weinkam, and Peter G Wolynes. Origins of barriers and barrierless folding in BBL. *Proceedings of the National Academy of Sciences of the United States of America*, 105(1):118–23, January 2008.

[182] Fang Huang, Liming Ying, and Alan R Fersht. Direct observation of barrier-limited folding of BBL by single-molecule fluorescence resonance energy transfer. *Proceedings of the National Academy of Sciences of the United States of America*, 106(38):16239–44, September 2009.

[183] G Horn, R Hofweber, W Kremer, and H R Kalbitzer. Structure and function of bacterial cold shock proteins. *Cell Mol Life Sci*, 64(12):1457–1470, 2007.

[184] W Kremer, B Schuler, S Harrieder, M Geyer, W Gronwald, C Welker, R Jaenicke, and H R Kalbitzer. Solution NMR structure of the cold-shock protein from the hyperthermophilic bacterium Thermotoga maritima. *Eur J Biochem*, 268(9):2527–2539, 2001.

[185] D Wassenberg, C Welker, and R Jaenicke. Thermodynamics of the unfolding of the cold-shock protein from Thermotoga maritima. *J Mol Biol*, 289(1):187–193, 1999.

[186] D Perl, C Welker, T Schindler, K Schroder, M A Marahiel, R Jaenicke, and F X Schmid. Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nat Struct Biol*, 5(3):229–235, 1998.

[187] Katarzyna M. Tych, Toni Hoffmann, David J. Brockwell, and Lorna Dougan. Single molecule force spectroscopy reveals the temperature-dependent robustness and malleability of a hyperthermophilic protein. *Soft Matter*, 9(37):9016, August 2013.

[188] Lia Cardarelli, Lisa G Pell, Philipp Neudecker, Nawaz Pirani, Amanda Liu, Lindsay A Baker, John L Rubinstein, Karen L Maxwell, and Alan R Davidson. Phages have adapted the same protein fold to fulfill multiple functions in virion assembly. *Proceedings of the National Academy of Sciences of the United States of America*, 107(32):14384–9, August 2010.

[189] Helios Murialdo, Xuekun Xing, Dimitra Tzamtzis, Abraham Haddad, and Marvin Gold. The product of the bacteriophage lambda W gene: purification and properties. *Biochemistry and cell biology = Biochimie et biologie cellulaire*, 81(4):307–15, August 2003.

[190] Lorenzo Sborgi, Abhinav Verma, Victor Muñoz, and Eva de Alba. Revisiting the NMR structure of the ultrafast downhill folding protein gpW from bacteriophage λ. *PloS one*, 6(11):e26409, January 2011.

[191] K L Maxwell, A R Davidson, H Murialdo, and M Gold. Thermodynamic and functional characterization of protein W from bacteriophage lambda. The three C-terminal residues are critical for activity. *The Journal of biological chemistry*, 275(25):18879–86, June 2000.

[192] Lorenzo Sborgi. *Analysis of protein-folding and protein-protein interactions at atomic resolution by NMR-spectroscopy*. PhD thesis, 2012.

[193] A Steward, J L Toca-Herrera, and J Clarke. Versatile cloning system for construction of multimeric proteins for use in atomic force microscopy. *Protein Sci*, 11(9):2179–2183, 2002.

[194] Javier Oroz, Rubén Hervás, Alejandro Valbuena, and Mariano Carrión-Vázquez. Unequivocal single-molecule force spectroscopy of intrinsically disordered proteins. *Methods in molecular biology (Clifton, N.J.)*, 896:71–87, January 2012.

[195] Bruker-Veeco. *di MultiMode PicoForce Instruction Manual*. 2004.

[196] Anil K. Jain. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8):651–666, June 2010.

[197] Daniel T. Gillespie. *Markov Processes: An Introduction for Physical Scientists*. 1991.

[198] Ravishankar Ramanathan and Victor Munoz. A method for extracting the free energy surface and conformational dynamics of fast-folding proteins from single molecule photon trajectories. *The journal of physical chemistry. B*, May 2015.

[199] J C Phillips, R Braun, W Wang, J Gumbart, E Tajkhorshid, E Villa, C Chipot, R D Skeel, L Kale, and K Schulten. Scalable molecular dynamics with NAMD. *J Comput Chem*, 26(16):1781–1802, 2005.

[200] B R Brooks, C L Brooks 3rd, A D Mackerell Jr., L Nilsson, R J Petrella, B Roux, Y Won, G Archontis, C Bartels, S Boresch, A Caflisch, L Caves, Q Cui, A R Dinner, M Feig, S Fischer, J Gao, M Hodoscek, W Im, K Kuczera, T Lazaridis, J Ma, V Ovchinnikov, E Paci, R W Pastor, C B Post, J Z Pu, M Schaefer, B Tidor, R M Venable, H L Woodcock, X Wu, W Yang, D M York, and M Karplus. CHARMM: the biomolecular simulation program. *J Comput Chem*, 30(10):1545–1614, 2009.

[201] W Humphrey, A Dalke, and K Schulten. VMD: visual molecular dynamics. *J Mol Graph*, 14(1):27–28,33–38, 1996.

[202] Hannes Neuweiler, Timothy D Sharpe, Trevor J Rutherford, Christopher M Johnson, Mark D Allen, Neil Ferguson, and Alan R Fersht. The folding mechanism of BBL: Plasticity of transition-state structure observed within an ultrafast folding protein family. *Journal of molecular biology*, 390(5):1060–73, July 2009.

[203] Emily J Guinn, Bharat Jagannathan, and Susan Marqusee. Single-molecule chemo-mechanical unfolding reveals multiple transition state barriers in a small single-domain protein. *Nature communications*, 6:6861, January 2015.